

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/57 C12N9/64 A61K38/48 G01N33/50 C12Q1/68  
C12N5/10 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A61K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	CELL, APR 7 1995, 81 (1) P27-40, UNITED STATES, XP002010548 RICHARD I ET AL: "Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A." see the whole document	1-20
X	--- JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 264, no. 33, 1989, BALTIMORE, MD US, pages 20106-20111, XP002010549 HIROYUKI SORIMACHI ET AL.: "Molecular cloning of a novel mammalian calcium-dependent protease distinct form both m- and mu-types" cited in the application see figures 2,3 --- -/-	1-8, 12-15

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

14 August 1996

Date of mailing of the international search report

20.08.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Espen, J

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 268, no. 26, 1993, BALTIMORE, MD US, pages 19476-19482, XP002010550 HIROYUKI SORIMACHI ET AL.: "A novel tissue-specific calpain species expressed predominantly in the stomach comprises two alternative splicing products with and without Ca <sup>2+</sup> -binding domain" see figure 2 ---	1-8, 12-15
X	BIOCHEMISTRY, vol. 27, 1988, EASTON, PA US, pages 8122-8128, XP002010551 SHINOBU IMAJOH ET AL.: "Molecular cloning of the cDNA for the large subunit of the High-Ca <sup>2+</sup> -Requiring form of human Ca <sup>2+</sup> -activated neutral protease" cited in the application see figure 2 ---	1-8, 12-15
Y	HUMAN MOLECULAR GENETICS, vol. 3, no. 2, February 1994, pages 285-293, XP002010552 FOUGEROUSSE F. ET AL.: "Mapping of a chromosome 15 region involved in limb girdle muscular dystrophy" cited in the application see the whole document ---	1-20
Y	JOURNAL OF MEDICAL GENETICS, vol. 30, 1993, pages 385-387, XP000578358 PASSOS-BUENO M. R. ET AL.: "Evidence of genetic heterogeneity in the autosomal recessive adult forms of limb-girdle muscular dystrophy following linkage analysis with 15q probes in Brazilian families" see the whole document ---	1-20
Y	GENOMICS, OCT 1994, 23 (3) P619-27, UNITED STATES, XP000578720 RICHARD I ET AL: "Regional localization of human chromosome 15 loci." see page 625 ---	1-20
A	INT. J. BIOCHEM., vol. 22, no. 8, 1990, pages 811-822, XP002010554 JOHNSON P. ET AL.: "Calpains (intracellular calcium-activated cysteine proteinases): structure-activity relationships and involvement in normal and abnormal cellular metabolism" see the whole document ---	
	---	

-/--

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p data-bbox="321 216 971 430">FASEB JOURNAL, vol. 8, no. 11, 1994, pages 814-822, XP002010555 TAKAOMI C. SAIDO ET AL.: "Calpain: new perspectives in molecular diversity and physiological-pathological involvement" see the whole document -----</p>	

# PCT

## INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference <b>B2628A - FL</b>	<b>FOR FURTHER ACTION</b> See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. <b>PCT/EP 95/ 04575</b>	International filing date (day/month/year) <b>21/11/1995</b>	Priority date (day/month/year) <b>22/11/1994</b>
International Patent Classification (IPC) or national classification and IPC <b>C12N15/57</b>		
Applicant <b>ASSOCIATION FRANCAISE CONTRE LES MYOPATHIES et al.</b>		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.


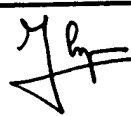
2. This REPORT consists of a total of 5 sheets, including this cover sheet.

☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consists of a total of \_\_\_\_\_ sheets.

3. This report contains indications and corresponding pages relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☒ Certain observations on the international application

Date of submission of the demand <b>19/06/1996</b>	Date of completion of this report <b>17. 03. 97</b>
Name and mailing address of the IPEA  European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk - Netherlands Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3018	Authorized officer  <b>J. Espen</b> Telephone No. <b>+31/70/340-2625</b>

# INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/EP95/04575

## I. Basis of the report

1. This report has been drawn up on the basis of *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.)*

☒ the international application as originally filed

☐ the description, pages \_\_\_\_\_, as originally filed  
 \_\_\_\_\_, filed with the demand  
 \_\_\_\_\_, filed with the letter of

☐ the claims, Nos. \_\_\_\_\_, as originally filed  
 \_\_\_\_\_, as amended under Article 19  
 \_\_\_\_\_, filed with the demand  
 \_\_\_\_\_, filed with the letter of

☐ the drawings, sheets / fig. \_\_\_\_\_, as originally filed  
 \_\_\_\_\_, filed with the demand  
 \_\_\_\_\_, filed with the letter of

2. The amendments have resulted in the cancellation of:

☐ the description, pages:

☐ the claims, Nos.

☐ the drawings, sheets / fig.

3. ☐ This opinion has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2 (c)).

4. Additional observations, if necessary:

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

## 1. Statement

Novelty	Claims	9-20	YES
	Claims	1-8	NO
Inventive Step	Claims	9-11,15-20	YES
	Claims	1-8,12-14	NO
Industrial Applicability	Claims	1-20	YES
	Claims		NO

## 2. Citations and Explanations

1). The present international application relates to Limb-Girdle Muscular Dystrophy Type 2A (LGMD2A). LGMD2A is strongly correlated to mutations in the gene of the proteolytic enzyme calpain 3 (*CANP3*).

The genomic organization of the human *CANP3* gene was determined.  
The above mentioned enzyme may also be referred to as 'p94' or 'nCL-1'.

2). The following document has been considered for the purposes of this report:

**D1** THE JOURNAL OF BIOLOGICAL CHEMISTRY, Vol. 264, pp. 20106-20111, 1989

3). The present application does not satisfy the criterion set forth in Article 33(2) PCT because the subject-matter of claims 1-8 is not new in respect of prior art as defined in the regulations (Rule 64(1)-(3) PCT).

Document D1 discloses the nucleotide sequence of the human *CANP3* gene and the deduced amino acid sequence (D1, Fig. 2).

In view of the above comment, D1 is novelty destroying for claims 1-8 (Art. 33 (2) PCT).  
The involvement of the *CANP3* gene in the etiology of LGMD2 is an implicit feature of the above sequences, and does not render novel the claimed subject-matter.

4). The methods of claims 12-14 are novel (Art. 33 (2) PCT), however they do not involve an inventive step (Art. 33 (3) PCT) for the following reasons.

**INTERNATIONAL PRELIMINARY EXAMINATION REPORT**

International application No.

**PCT/EP95/04575**

For the person skilled in the art, being aware of the known amino acid sequences of claims 5 to 7 or of the known nucleic acid sequences of claims 1-4, the carrying out of said methods required nothing out of the ordinary and thus involved no inventive skill, all being a matter of technical convenience. In addition, this was a matter of normal design procedures for which neither 'creative thinking' nor 'inventive talent' were necessary.

5). Claims 9-11, and 15-20 do meet the requirements of Art. 33 (2) and (3) for the following reasons:

Before the priority date of the present international application, it was neither disclosed nor suggested that the *CANP3* gene product, when mutated, is involved in the LGMD2 disease.

In consequence, the use of sequences related to the *CANP3* gene in the diagnostic of the LGMD2 disease or their use in the preparation of a pharmaceutical composition for the treatment of said disease could not be deduced in an obvious manner from the closest prior art document D1.

The industrial applicability of the present set of claims is acknowledged (Art. 33 (4) PCT).

# INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/EP95/04575

## VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

- Claim 6 should read '..according to claim 5, ...'.
- Claim 18 should read '...by nucleic acid amplification...'



## PCT

## INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference <b>B2628A - FL</b>	<b>FOR FURTHER ACTION</b> see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. <b>PCT/EP 95/04575</b>	International filing date (day/month/year) <b>21/11/95</b>	(Earliest) Priority Date (day/month/year) <b>22/11/94</b>
Applicant <b>ASSOCIATION FRANCAISE CONTRE LES MYOPATHIES et al.</b>		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 4 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. ☐ Certain claims were found unsearchable (see Box I).

2. ☐ Unity of invention is lacking (see Box II).

3. ☒ The international application contains disclosure of a nucleotide and/or amino acid sequence listing and the international search was carried out on the basis of the sequence listing

☐ filed with the international application.

☒ furnished by the applicant separately from the international application,

☐ but not accompanied by a statement to the effect that it did not include matter going beyond the disclosure in the international application as filed.

☐ Transcribed by this Authority

4. With regard to the title, ☐ the text is approved as submitted by the applicant.

☒ the text has been established by this Authority to read as follows:

**LGMD GENE CODING FOR A CALCIUM DEPENDENT PROTEASE**

5. With regard to the abstract,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the drawings to be published with the abstract is:

Figure No. \_\_\_\_\_ ☐ as suggested by the applicant.

☐ because the applicant failed to suggest a figure.

☐ because this figure better characterizes the invention.

☒ None of the figures.

## INTERNATIONAL SEARCH REPORT

National Application No

PCT/EP 95/04575

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/57 C12N9/64 A61K38/48 G01N33/50 C12Q1/68  
C12N5/10 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A61K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	CELL, APR 7 1995, 81 (1) P27-40, UNITED STATES, XP002010548 RICHARD I ET AL: "Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A." see the whole document	1-20
X	--- JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 264, no. 33, 1989, BALTIMORE, MD US, pages 20106-20111, XP002010549 HIROYUKI SORIMACHI ET AL.: "Molecular cloning of a novel mammalian calcium-dependent protease distinct from both m- and mu-types" cited in the application see figures 2,3 --- -/--	1-8, 12-15

☒ Further documents are listed in the continuation of box C.☐ Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

14 August 1996

Date of mailing of the international search report

20.08.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Espen, J

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 268, no. 26, 1993, BALTIMORE, MD US, pages 19476-19482, XP002010550 HIROYUKI SORIMACHI ET AL.: "A novel tissue-specific calpain species expressed predominantly in the stomach comprises two alternative splicing products with and without Ca <sup>2+</sup> -binding domain" see figure 2 ---	1-8, 12-15
X	BIOCHEMISTRY, vol. 27, 1988, EASTON, PA US, pages 8122-8128, XP002010551 SHINOBU IMAJOH ET AL.: "Molecular cloning of the cDNA for the large subunit of the High-Ca <sup>2+</sup> -Requiring form of human Ca <sup>2+</sup> -activated neutral protease" cited in the application see figure 2 ---	1-8, 12-15
Y	HUMAN MOLECULAR GENETICS, vol. 3, no. 2, February 1994, pages 285-293, XP002010552 FOUGEROUSSE F. ET AL.: "Mapping of a chromosome 15 region involved in limb girdle muscular dystrophy" cited in the application see the whole document ---	1-20
Y	JOURNAL OF MEDICAL GENETICS, vol. 30, 1993, pages 385-387, XP000578358 PASSOS-BUENO M. R. ET AL.: "Evidence of genetic heterogeneity in the autosomal recessive adult forms of limb-girdle muscular dystrophy following linkage analysis with 15q probes in Brazilian families" see the whole document ---	1-20
Y	GENOMICS, OCT 1994, 23 (3) P619-27, UNITED STATES, XP000578720 RICHARD I ET AL: "Regional localization of human chromosome 15 loci." see page 625 ---	1-20
A	INT. J. BIOCHEM., vol. 22, no. 8, 1990, pages 811-822, XP002010554 JOHNSON P. ET AL.: "Calpains (intracellular calcium-activated cysteine proteinases): structure-activity relationships and involvement in normal and abnormal cellular metabolism" see the whole document ---	
	--- -/--	

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 95/04575

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>FASEB JOURNAL, vol. 8, no. 11, 1994, pages 814-822, XP002010555 TAKAOMI C. SAIDO ET AL.: "Calpain: new perspectives in molecular diversity and physiological-pathological involvement" see the whole document -----</p>	



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/57, 9/64, A61K 38/48, G01N 33/50, C12Q 1/68, C12N 5/10, A61K 48/00</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 96/16175</b> <b>(43) International Publication Date:</b> 30 May 1996 (30.05.96)
<b>(21) International Application Number:</b> PCT/EP95/04575 <b>(22) International Filing Date:</b> 21 November 1995 (21.11.95)  <b>(30) Priority Data:</b> 94402668.1 22 November 1994 (22.11.94) EP <b>(34) Countries for which the regional or international application was filed:</b> GB et al.  <b>(71) Applicant (for all designated States except US):</b> ASSOCIATION FRANÇAISE CONTRE LES MYOPATHIES [FR/FR]; 13, place de Rungis, F-75013 Paris (FR).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> BECKMANN, Jacques [FR/FR]; 95, rue de Paris, F-94220 Charenton-le-Pont (FR). RICHARD, Isabelle [FR/FR]; 72, rue de l'Essonne, F-91000 Evry (FR).  <b>(74) Agents:</b> GUTMANN, Ernest et al.; Ernest Gutmann - Yves Plasseraud S.A., 3, rue Chauveau-Lagarde, F-75008 Paris (FR).		<b>(81) Designated States:</b> CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> LGMD GENE CODING FOR A CALCIUM DEPENDENT PROTEASE  <b>(57) Abstract</b>  A nucleic acid sequence comprising: 1) the sequence represented in figure 8; or 2) the sequence represented in figure 2; or 3) a part of the sequence of figure 2 with the proviso that it is able to code for a protein having a calcium dependant protease activity involved in a LGMD2; or 4) a sequence derived from a sequence defined in 1), 2) or 3) by substitution, deletion or addition of one or more nucleotides with the proviso that said sequences still codes for said protease.		



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :C12N 15/57, 9/64, A61K 38/48, G01N  
33/50, C12Q 1/68, C12N 5/10, A61K  
48/00

A2

(11) International Publication Number:

WO 96/16175

(43) International Publication Date:

30 May 1996 (30.05.96)

(21) International Application Number: PCT/EP95/04575

(22) International Filing Date: 21 November 1995 (21.11.95)

(30) Priority Data:

94402668.1 ✓ 22 November 1994 (22.11.94) EP

(34) Countries for which the regional or  
international application was filed: GB et al.(71) Applicant (for all designated States except US): ASSOCI-  
ATION FRANÇAISE CONTRE LES MYOPATHIES  
[FR/FR]; 13, place de Rungis, F-75013 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BECKMANN, Jacques  
[FR/FR]; 95, rue de Paris, F-94220 Charenton-le-Pont (FR).  
RICHARD, Isabelle [FR/FR]; 72, rue de l'Essonne, F-91000  
Evry (FR).(74) Agents: GUTMANN, Ernest et al.; Ernest Gutmann - Yves  
Plasseraud S.A., 3, rue Chauveau-Lagarde, F-75008 Paris  
(FR).(81) Designated States: CA, JP, US, European patent (AT, BE, CH,  
DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

Without international search report and to be republished  
upon receipt of that report.

(54) Title: LGMD GENE CODING FOR A CALCIUM DEPENDENT PROTEASE

(57) Abstract

A nucleic acid sequence comprising: 1) the sequence represented in figure 8; or 2) the sequence represented in figure 2; or 3) a part of the sequence of figure 2 with the proviso that it is able to code for a protein having a calcium dependant protease activity involved in a LGMD2; or 4) a sequence derived from a sequence defined in 1), 2) or 3) by substitution, deletion or addition of one or more nucleotides with the proviso that said sequences still codes for said protease.

## LGMD gene coding for a calcium dependent protease

The invention relates to the isolated gene coding for a calcium dependent protease belonging to the Calpain family which, when it is mutated, is a cause of a disease called Limb-Girdle Muscular Dystrophy (LGMD).

The term limb-girdle muscular dystrophy (LGMD) was first proposed by Walton and Nattrass (1954) as part of a classification of muscular dystrophies. LGMD is characterised by progressive symmetrical atrophy and weakness of the proximal limb muscles and by elevated serum creatine kinase. Muscle biopsies demonstrate dystrophic lesions and electromyograms show myopathic features. The symptoms usually begin during the first two decades of life and the disease gradually worsens, often resulting in loss of walking ability 10 or 20 years after onset (Bushby, 1994). Yet, the precise nosological definition of LGMD still remains unclear. Consequently, various neuromuscular diseases such as facioscapulohumeral, Becker muscular dystrophies and especially spinal muscular atrophies have been occasionally classified under this diagnosis. For example, a recent study (Arikawa et al., 1991) reported that 17% (out of 41) of LGMD patients showed a dystrophinopathy. These issues highlight the difficulty in undertaking an analysis of the molecular and genetic defect(s) involved in this pathology.

Attempts to identify the genetic basis of this disease go back over 35 years. Morton and Chung (1959) estimated that "the frequency of heterozygous carrier ... is 16 per thousand persons". The same authors also stated that "the segregation analysis gives no evidence on whether these genes in different families are allelic or at different loci". Both autosomal dominant and recessive transmission have been reported, the latter being more common with an estimated prevalence of  $10^{-5}$  (Emery, 1991). The localisation of a gene for a recessive form on chromosome 15 (LGMD2A, MIM 253600; Beckmann et al., 1991) provided the definitive proof that LGMD is a specific genetic entity. Subsequent genetic analyses confirmed this chromosome 15 localisation (Young et al., 1992; Passos-Bueno et al., 1993), the latter group demonstrating genetic heterogeneity of this disease. Although a recent study localised a second mutant

gene to chromosome 2 (LGMD2B, MIM 253601; Bashir et al., 1994). there is evidence that at least one other locus can be involved.

Genetic analyses of the LGMD2 kindreds revealed unexpected findings. First genetic heterogeneity was demonstrated in the highly inbred Indiana Amish community. Second although the Isle of la Réunion families were thought to represent a genetic isolate, at least 6 different disease haplotypes were observed, providing evidence against the hypothesis of a single founder effect (Beckmann et al., 1991) in this inbred population.

The nonspecific nosological definition, the relatively low prevalence and genetic heterogeneity of this disorder limit the number of families which can be used to restrict the genetic boundaries of the LGMD2A interval. Cytogenetic abnormalities, which could have helped to focus on a particular region, have not been reported. Immunogenetic studies of dystrophin-associated proteins (Matsumura et al., 1993) and cytoskeletal or extracellular matrix proteins such as a merosin (Tomé et al., 1994) failed to demonstrate any deficiency. In addition, there is no known specific physiological feature or animal model that could help to identify a candidate gene. Thus, there is no alternative to a positional cloning strategy.

It is established that the LGMD2 chromosomal region is localized on chromosome 15 as 15q15.1 - 15q21.1 region (Fougerousse et al., 1994).

Construction and analysis of a 10-12 Mb YAC contig (Fougerousse et al., 1994) permitted the mapping of 33 polymorphic markers within this interval and to further narrow the LGMD2A region to between D15S514 and D15S222. Furthermore, extensive analysis of linkage disequilibrium suggested a likely position for the gene in the proximal part of the contig.

The invention results from the construction of a partial cosmid map and the screening by cDNA selection (Lovett et al., 1991 ; Tagle et al., 1993) for muscle-expressed sequences encoded by this interval led to the identification of a number of potential candidate genes. One of these, previously cloned by Sorimachi et al. (1989), encodes a muscle specific protein, nCL1 (novel Calpain Large subunit 1), which belongs to the calpain family (CANP, calcium-activated neutral protease; EC 3.4.22.17), and appeared to be a functional candidate gene for this disease.



Calpains are non-lysosomal intracellular cysteine proteases which require calcium for their catalytic activities (for a review see Croall D.E. et al, 1991). The mammalian calpains include two ubiquitous proteins CANP1 and CANP2 as well as tissue-specific proteins. In addition to the muscle specific nCL1, stomach specific nCL2 and nCL2' proteins have also been described; these are derived from the same gene by alternative splicing. The ubiquitous enzymes consist of heterodimers with distinct large subunits associated with an common small subunit ; the association of tissue-specific large subunits with a small subunit has not yet been demonstrated. The large subunits of calpains can be subdivided into 4 protein domains. Domains I and III, whose functions remain unknown, show no homology with known proteins. Domain I, however, seems important for the regulation of the proteolytic activity. Domain II shows similarity with other cysteine proteases, sharing histidine, cysteine and asparagine residues at its active sites. Domain IV comprises four EF-hand structures which are potential calcium binding sites. In addition, three unique regions with no known homology are present in the muscle-specific nCL1 protein, namely NS, IS1 and IS2, the latter containing a nuclear translocation signal. These regions may be important for the muscle specific function of nCL1.

It is usually accepted that muscular dystrophies are associated with excess or deregulated calpains, and all the known approaches for curing these diseases are the use of antagonists of these proteases ; examples are disclosed in EP 359309 or EP 525420.

The invention results from the finding that, on the opposite to all these hypothesis, the LGMD2 disease is strongly correlated to the defect of a calpain which is expressed in healthy people.

The invention relates to the nucleic acid sequence such as represented in Figure 2 coding for a  $\text{Ca}^{++}$  dependent protease, or calpain, which is involved in LGMD2 disease, and more precisely LGMD2A. It also relates to a part of this sequence provided it is able to code for a protein having a calcium-dependent protease activity involved in LGMD2, or a sequence derived from one of the above sequences by substitution, deletion or addition of one or more nucleotides provided that said sequence is still coding for said protein, all the nucleic acids yielding a sequence complementary to a sequence as defined above.

The genomic organisation of the human nCL1 gene has been determined by the inventors, and consists of 24 exons and extends over 40 kb as represented in Figure 8, and is also a part of the invention. About 35 kb of this gene have been sequenced. A systematic screening of this gene in LGMD2A families led to the identification of 14 different mutations, establishing that a number of independent mutational events in nCL1 are responsible for LGMD2A. Furthermore, this is the first demonstration of a muscular dystrophy resulting from an enzymatic rather than a structural defect.

In the present specification, CANP3 means the protein which is a  $\text{Ca}^{++}$  dependent protease, or calpain, and coded by the nCL1 gene on chromosome 15.

The invention relates also to a protein, called CANP3, consisting in the amino acid sequence such as represented in figure 2 and which is involved, when mutated, in the LGMD2 disease.

The cDNA of the gene coding for CANP3, which is coding for the protein, is also represented in Figure 2, and is a part of the invention.

The protein coded by this DNA is CANP3, a calcium-dependent protease belonging to the Calpain family.

Are also included in the present invention the nucleic acid sequences derived from the cDNA of Figure 2 by one or more substitutions, deletions, insertions, or by mutations in 5' or 3' non coding regions or in splice sites, provided that the translated protein has the protease, calcium-dependent activity, and when mutated, induce LGMD2 disease.

The nucleic acid sequence encoding the protein might be DNA or RNA and be complementary to the nucleic acid sequence represented in Figure 2.

The invention also relates to a recombinant vector including a DNA sequence of the invention, under the control of a promoter allowing the expression of the calpain in an appropriate host cell.

A procaryotic or eucaryotic host cell transformed by or transfected with a DNA sequence comprising all or part of the sequence of Figure 2 is a part of the invention.

Such a host cell might be either :

- a cell which is able to secrete the protein and, this recombinant protein might be used as a drug to treat the LGMD2, or

- a packaging cell line transfected by a viral or retroviral vector ; the cell lines bearing recombinant vector might be used as a drug for gene therapy of LGMD2.

All the systems used today for gene therapy including adenoviruses and retroviruses and others described for example in « l'ADN médicament », (John Libbey, Eurotext, 1993), and bearing one of the DNA sequence of the invention are included herein by reference.

The examples hereunder and attached figures indicate how the structure of the gene was established, and how relationship between the gene and the LGMD was established.

Legend of the figures :

Figure 1:

A) Genomic organisation of the nCL1 gene

The gene covers a 40 kb region of which 35 were sequenced (Accession number pending). Introns and exons are drawn to scale, the latter being indicated by numbered vertical bars. The first intron is the largest one and remains to be fully sequenced. Position of intragenic microsatellites are indicated by asterisks. Arrows indicate the orientation of Alu (closed) and of Mer2 (greyed) repeat sequences.

B) *EcoRI* restriction map

An *EcoRI* (E) restriction map of this region was established with the help of cosmids from this region. The location of nCL1 gene is indicated as a black bar. The size of the corresponding fragments are indicated and are underlined when determined by sequence analysis.

C) Cosmid map of the nCL1 gene region.

Cosmids were from a cosmid library constructed by subcloning YAC 774G4 (Richard in preparation) and are presented as lines. Dots on lines indicate positive STSs (indicated in boxed rectangles). A minimum of three cosmids cover the entire gene. T3, T7

**Figure 2:** Sequence of the human nCL1 cDNA (B) , and the flanking 5' (A) and 3' (C) genomic regions.

A) and C) The polyadenylation signal and putative CAAT, TATAA sites are boxed. Putative Sp1 (position -477 to -472), MEF2 binding sites (-364 to -343) and CArG box (-685 to -672) are in bold. The Alu sequence present in the 5' region is underlined.

B) The corresponding amino acids are shown below the sequence. The coding sequence between the ATG initiation codon and the TGA stop codon is 2466 bp, encoding for a 821 amino acid protein. The adenine in the first methionine codon has been assigned position 1. Locations of introns within the nCL1 gene are indicated by arrowheads. Nucleotides which differ from the previously published ones are indicated by asterisks.

**Figure 3:** Alignments of amino acid sequences of the muscle-specific calpains.

The human nCL1 protein is shown on the first line. The 3 muscle-specific sequences (NS, IS1 and IS2) are underlined. The second line corresponds to the rat sequence (Accession no P). The third and fourth lines show the deduced amino acid sequences encoded by pig and bovine Expressed Sequences Tagged (GenBank accession no U05678 and no U07858, respectively). The amino acids residues which are conserved among all known members of the calpains are in reverse letters. A period indicates that the same amino acid is present in the sequence. Letters refer to the variant amino acid found in the homologous sequence. Position of missense mutations are given as numbers above the mutated amino acid.

**Figure 4:** Distribution of the mutations along nCL1 protein structure.

A) Positions of the 23 introns are indicated by vertical bars in relation to the corresponding amino acid coordinates.

B) The nCL1 protein is depicted showing the four domains (I, II, III, IV) and the muscle specific sequences (NS, IS1 and IS2). The position of missense mutations within nCL1 domain are indicated by black dots. The effect of nonsense and frameshift mutations are illustrated as truncated lines, representing the extent of protein synthesised. Name of the corresponding families are indicated on the left of the line. The out of frame ORF is given by hatched lines.

Figure 5: Northern blot hybridisation of a nCL1 clone

A mRNA blot (Clontech) containing 2 µg of poly(A)+ RNA from each of eight human tissues was hybridised with a nCL1 genomic clone spanning exons 20 and 21. The latter detects a 3.6 kb mRNA present only in a line corresponding to the skeletal muscle mRNA.

Figure 6: Representative mutations identified by heteroduplex analysis.

Examples of mutation screening by heteroduplex analysis. Pedigree B505 shows the segregation of two different mutations in exon 22.

Figure 7: Homozygous mutations in the nCL1 gene

Detection by sequencing of mutations in exons 2 (a), 8 (b), 13 (c) and 22 (d). Sequences from a healthy control are shown above each mutant sequence. Asterisks indicate the position of the mutated nucleotides. The consequences on codon and amino acid residues are indicated on the left of the figure together with the name of the family.

Figure 8 : Structure of nCL1 gene

Figure 8A represents the 5' part of the gene with exon 1.

Figure 8B represents the part of the gene including exons 2 to 8,

Figure 8C represents the part of the gene including exon 9,

Figure 8D represents the part of the gene including exons 10 to 24 including the 3' non transcribed region.

**EXAMPLES**

**EXAMPLE 1**

**Localisation of the nCL1 within the LGMD2A interval**

Detailed genetic and physical maps of the LGMD2A region were constructed (Fougerousse et al., 1994), following the primary linkage assignment to 15q (Beckmann et al., 1991). The disease locus was bracketed between the D15S129 and D15S143 markers, defining the cytogenetic boundaries of the LGMD2A region as 15q15.1-15q21.1 (Fougerousse et al., 1994). Construction and analysis of a 10-12 Mb YAC contig (Fougerousse et al., 1994) permitted us to map 33 polymorphic markers within this interval and to further narrow the LGMD2A region to between D15S514 and D15S222.

The nCL1 gene had been localised to chromosome 15 by hybridisation with sorted chromosomes and by Southern hybridisation to DNA from human-mouse cell hybrids (Ohno et al., 1989). cDNA capture using YACs from the LGMD2A interval allowed the identification of thirteen positional candidate genes. nCL1 was one of the two transcripts identified that showed muscle-specific expression as evidenced by northern blot analysis. The localisation was further confirmed by STS (for Sequence Tagged Site) assays. Primers used for the localisation of the nCL1 gene are P94in2, P94in13 and pcr6a3, as shown in Figure 1 and their characteristics being defined in Table 1.

Table 1: PCR primers used for localisation of the nCL1 gene.

Primer name	Primer sequence (5'-3')	Position within the cDNA	Annealing temp (°C)	PCR product size on	
				cDNA	genomic DNA
P94in2	ATGGAGCCAACAGAACTGA C GTATGACTCGGAAAAGAAG GT	341-360 428-448	58	108	1758
P94in13	TAAGCAAAAGCAGTCCCCA C TTGCTGTTCTCACTTTCCT G	1893-1912 1936-1956	58	64	1043
P94-6a3	GTTTCATCTGCTGCTTCGTT CTGGTTCAGGCATACATGG T	2342-2361 2452-2471	56	130	818
P94ex1ter	TTCTTTATGTGGACCCTGAG TT ACGAACTGGATGGGGAACT	218-239 275-293	55	76	76

These primers are designed from different parts of the published human cDNA sequence (Sorimachi et al., 1989), and were used for an STS content screening on DNA from three chromosome 15 somatic cell hybrids and YACs from the LGMD2A contig. The results positioned the gene in a region previously defined as 15q15.1-q21.1 and on 3 YACs (774G4, 926G10, 923G7) localised in this region. The relative positions of STSs along the LGMD2A contig allowed to localise the gene between D15S512 and D15S488, in a candidate region suggested by linkage disequilibrium studies.

The same primers as above were used to screen a cosmid library from YAC 774G4. A group of 5 cosmids was identified (Fig. 1). Experiments with another nCL1 primer pair (P94ex1ter; Table 1) established that these cosmids cover all nCL1 exons except number 1, and that a second group of 4 cosmids contain this

exon (Fig. 1). A minimal set of three overlapping cosmids (2G8-2B11-1F11) covers the entire gene (Figure 1). DNA from these cosmids was used to construct an *EcoRI* restriction map of this region (Figure 1B).

## EXAMPLE 2

### 5 **Determination of the nCL1 gene sequence**

Most of the sequences were obtained through shotgun sequencing of partial digests of cosmid 1F11 subcloned in M13 and bluescript vectors, and by walking with internal primers. The sequence assembly was made using the XBAP software of the Staden package (Staden) and was in agreement with the  
10 restriction map of the cosmids. Sequences of exon 1 and adjacent regions were obtained by sequencing cosmid DNA or PCR products from human genomic DNA. The first intron is still not fully sequenced, but there is evidence that it may be between 10 to 16 kb in length (based on hybridisation of restriction fragments; data not shown). The entire gene, including its 5' and 3' regions, is more than 40  
15 kb long, and shown in Figure 8.

#### a) the cDNA sequence

The used technology allows the implementation of the published human cDNA sequence of nCL1 (Sorimachi 1989). It contains the missing 129 bases corresponding to the N-terminal 43 amino acids (Figure 2). It also differs from it  
20 at 12 positions. Three of which occur at third base positions of codons and preserve the encoded amino acid sequence. The other 9 differences lead to changes in amino-acid composition (Figure 2). As these different exons were sequenced repeatedly on at least 10 distinct genomes, we are confident that the sequence of Fig. 2 represents an authentic sequence and does not contain  
25 minor polymorphic variants. Furthermore, these modifications increase the local similarity with the rat nCL1 amino acid sequence (Sorimachi), although the overall similarity is still 94 %.

The ATG numbered 1 in Figure 2 is the translation initiation site based on homology with the rat nCL1, and is within a sequence with only 5 nucleotides out  
30 of 8 in common with the Kosak consensus sequence (Kosak M, 1984). Putative CCAAT and TATA boxes were observed 590, 324, (CCAAT) and 544 or 33 bp (TATA) upstream of the initiating ATG codon, respectively (Bucher, 1990). A GC-box binding the Sp1 protein (Dyran et al., 1983) was identified at position -477.

Consensus sequences corresponding to potential muscle-specific regulatory elements were identified (Fig. 2). These include a myocyte-specific enhancer-binding factor 2 (MEF2) binding site (Cserjesi P. 1991), a CArG box (Minty A. 1986) and 6 E-boxes (binding sites for basic Helix-Loop-Helix proteins frequently found in members of MyoD family; Blackwell et Weintraub, 1990). The functional significance of these putative transcription factor binding sites in the regulation of nCL1 gene expression remains to be established.

Two potential AAUAAA polyadenylation signals, were identified 520 and 777 bp downstream of the TGA stop codon. The sequencing of a partial nCL1 cDNA containing a polyA tail, demonstrated that the first AAUAAA is the polyadenylation signal. The latter is embedded in a region well conserved with the rat nCL1 sequence and is followed after 4 bp by a G/T cluster, present in most genes 3' of the polyadenylation site (Birnstiel et al., 1985). The 3'-untranslated region of the nCL1 mRNA is 565 bp long. The predicted length of the cDNA should therefore be approximately 3550 or 3000 bp.

#### b) Comparison with calpain

The sequence of the human nCL1 gene was compared to those of other calpains thereof (Figure 3). The most telling comparisons are with the homologous rat (Accession no J05121), bovine (Accession no U07858) and porcine (Accession no U05678) sequences. The accession numbers refers to those or international genebanks, such as GeneBank (N.I.H.) or EMBL Database (EMBL, Heidelberg). High local similarities between the human and rat DNA sequences are even observed in the 5' (75%) or in different parts of the 3' untranslated regions (over 60%) (data not shown). The high extent of sequence homology manifested by the human and rat nCL1 gene in their untranslated regions is suggestive of evolutionary pressures on common putative regulatory sequences.

#### c) Genomic organisation of the nCL1 gene

A comparison of the published nCL1 human cDNA (Sorimachi et al., 1989) with the corresponding genomic sequence led to the identification of 24 exons ranging in length from 12 bp (exon 13) to 309 bp (exon 1), with a mean size of 100 bp (Figure 1). The size of introns ranges from 86 bp to about 10-16 kb for intron 1.



The intron-exon boundaries as shown in Table 2 exhibit close adherence to 5' and 3' splice site consensus sequences (Shapiro and Senapathy, 1987).

**Table 2:** Sequences at the intron-exon junctions. A score expressing adherence to the consensus was calculated for each site according to Shapiro and Senapathy (1987). Sequences of exons and introns are in upper and lower cases, respectively. Size of exons are given in parenthesis.

splice donor site	score (%)	Intron	score (%)	splice acceptor site	Exon
					Exon 1 (309 bp) ->
...CTCCGgtgagt...	88.5	<-Intron 1->	99.0	...tttgttcacagGAAAT...	Exon 2 (70 bp) ->
...GCTAGgttagga...	83.5	<-Intron 2->	90.0	...gtgtctgcctgcagGGGAC...	Exon 3 (119 bp) ->
...TCCAGgtgagg...	92	<-Intron 3->	81.5	...acgcttctgtgcagTTCTG...	Exon 4 (134 bp) ->
...GCTAAGtaagc...	82	<-Intron 4->	81.5	...atcctctcttaagGCTCC...	Exon 5 (169 bp) ->
...TTGATgtaagt...	87	<-Intron 5->	79.5	...ccatcgggcctcagGATGG...	Exon 6 (144 bp) ->
...CCCGGgtgigt...	77.5	<-Intron 6->	91	...ttactgtcttacagACAAT...	Exon 7 (84 bp) ->
...ATGAGgttaagc...	94	<-Intron 7->	78.5	...tctgtgtgcttaagGTCCC...	Exon 8 (86 bp) ->
...GATAGgttaggt...	89	<-Intron 8->	91.5	...cattttcccaccagATGGA...	Exon 9 (78 bp) ->
...TTCTGgtgagt...	88	<-Intron 9->	92	...ttccaacctctcagGATGT...	Exon 10 (161 bp) ->
...CCCAGgtggga...	80	<-Intron 10->	68.5	...ttctgggggtgcagATACT...	Exon 11 (170 bp) ->
...ACGAGgtgigt...	85.5	<-Intron 11->	86	...tgttcttctcaagGTTCC...	Exon 12 (12 bp) ->
...AAGAGgtatag...	70	<-Intron 12->	87	...tccccatctctcagATGCA...	Exon 13 (209 bp) ->
...TCTGAgtagt...	76.5	<-Intron 13->	97	...tgtattcctcacagGGAAG...	Exon 14 (37 bp) ->
...CAGTGgtgagt...	89	<-Intron 14->	93.5	...cttttctatgcagAAAAA...	Exon 15 (18 bp) ->
...CCAAGgttaggt...	89	<-Intron 15->	87	...cctcctctctccagCCCAT...	Exon 16 (114 bp) ->
...CACAGgtgict...	80	<-Intron 16->	88	...ttgtgcctccacagCCACA...	Exon 17 (78 bp) ->
...GAGATgtgagt...	84	<-Intron 17->	92.5	...cccttctcctcagGACAT...	Exon 18 (58 bp) ->
...CAAACgtgagt...	83	<-Intron 18->	90	...ctccatccccccagACAAG...	Exon 19 (65 bp) ->
...TGGATgtatcc...	56	<-Intron 19->	88	...cctccctcctccagACAGA...	Exon 20 (69 bp) ->
...GGCAGgtggga...	80	<-Intron 20->	94	...ttttctattgccagAAATA...	Exon 21 (79 bp) ->
...CGCAGgtgcig...	66	<-Intron 21->	91	...gggccctccacagGATTC...	Exon 22 (117 bp) ->

...GTTCAgtaagi...	79	<-Intron 22->	93.5	...gcattctttcacagGAGCT...	Exon 23 (59 bp) ->
...TGGAGgtaaag...	81	<-Intron 23->	79	...gggaacttcttcagTGGCT...	Exon 24 (27 bp) ->

When the genomic sequence was submitted to GRAIL analysis (Uberbacher et al., 1991), 11 exons were correctly recognised, 4 were not identified, 6 were inadequately defined and 2 were too small to be recognised (data not shown).

5 As already noted, the nCL1 gene has three unique sequence blocks, NS (amino acid residues 1 to 61), IS1 (residues 267 to 329) and IS2 (residues 578 to 653). It is interesting to note that each of these sequences, as well as the nuclear translocation signal inside IS2, are essentially flanked by introns (Fig. 4). The exon-intron organisation of the human nCL1 is similar to that reported for  
10 the chicken CANP (the only other large subunit calpain gene whose genomic structure is known; (Emori et al., 1986).

Four microsatellite sequences were identified. Two of them are in the distal part of the first intron: an (AT)<sub>14</sub> and an previously identified mixed-pattern microsatellite, S774G4B8, which was demonstrated to be non polymorphic  
15 (Fougerousse et al., 1994). A (TA)<sub>7</sub>(CA)<sub>4</sub>(GA)<sub>13</sub> was identified in the second intron and genotyping of 64 CEPH unrelated individuals revealed two alleles (with frequencies of 0.10 and 0.90). The fourth microsatellite is a mixed (CA)<sub>n</sub>(TA)<sub>m</sub> repeat present in the 9th intron. The latter and the (AT)<sub>14</sub> repeat have not been investigated for polymorphism. Fourteen repetitive sequences of  
20 the Alu family and one Mer2 repeat were identified in the nCL1 gene (Fig. 1C), which has, thus, on the average one Alu element per 2.5 kb.

Southern blot experiments (Ohno et al., 1989) and STS screening (data not shown) suggest that there is but one copy per genome of this member of the calpain family.

### 25 EXAMPLE 3

#### Expression of the nCL1 gene

The pattern of tissue-specificity was investigated by northern blot hybridisation with a genomic subclone probe from cosmid 1F11 spanning exons 20 and 21. There is no evidence for the existence of an alternatively spliced form  
30 of nCL1, although this cannot be excluded. A transcript of about 3.4-3.6 kb was

detected in skeletal muscle mRNA (Figure 5). This size therefore favours that the position -544 is the functional TATA box.

Transcription studies suggested that it is an active gene rather than a pseudogene and its muscle-specific pattern of expression is consistent with the phenotype of this disorder (Sorimachi et al., 1989 and Figure 5).

#### EXAMPLE 4

##### Mutation screening

nCL1 fulfils both positional and functional criteria to be a candidate gene for LGMD2A. To evaluate its role in the etiology of this disorder, nCL1 was systematically screened in 38 LGMD2 families for the presence of nucleotide changes using a combination of heteroduplex (Keen et al., 1991) and direct sequence analyses.

PCR primers were designed to specifically amplify the exons and splice junctions and also the regions containing the putative CAT, TATA boxes and the polyadenylation signal of the gene as shown in Table 3.

**Table 3:** PCR primers used for the analysis of the nCL1 gene in LGMD patients.

amplified region	Primer sequences (5'-3')	Size (bp)	Annealing temp. (°C)
promotor	TTCAGTACCTCCCGTTCACC GATGCTTGAGCCAGGAAAAC	296	59
exon 1	CTTTCCTTGAAGGTAGCTGTAT GAGGTGCTGAGTGAGAGGAC	438	60
exon 2	ACTCCGTCTCAAAAAAATACCT ATTGTCCCTTTACCTCCTGG	239	57
exon 3	TGGAAGTAGGAGAGTGGGCA GGGTAGATGGGTGGGAAGTT	354	58
exon 4	GAGGAATGTGGAGGAAGGAC TTCCTGTGAGTGAGGTCTCG	292	59
exon 5	GGAACTCTGTGACCCCAAAT TCCTCAAACAAAACATTTCGC	325	56
exon 6	GTTCCCTACATTCTCCATCG GTTATTTCAACCCAGACCCTT	315	57
exon 7	AATGGGTTCTCTGGTTACTGC AGCACGAAAAGCAAAGATAAA	333	56
exon 8	GTAAGAGATTGCCCCCAG TCTGCGGATCATTGGTTTTG	321	58
exon 9	CCTTCCCTTCTTCCTGCTTC CTCTCTTCCCCACCCTTACC	173	56
exon 10	CCTCCTCACCTGCTCCCAT TTTTTCGGCTTAGACCCTCC	251	56
exon 11	TGTGGGGAATAGAAATAAATGG CCAGGAGCTCTGTGGGTCA	355	57
exon 12	GGCTCCTCATCCTCATTACA GTGGAGGAGGGTGAGTGTGC	312	61
exon 13	TGTGGCAGGACAGGACGTTC	337	60

	14		
exon 14	TTCAACCTCTGGAGTGGGCC	230	61
	CACCAGAGCAAACCGTCCAC		
exon 15	ACAGCCCAGACTCCCATTCC	225	57
	TTCTCTTCTCCCTTCACCCT		
exon 16	ACACACTTCATGCTCTCTACCC	331	56
	CCGCCTATTCCTTTCCTCTT		
exon 17	GACAAACTCCTGGGAAGCCT	270	61
	ACCTCTGACCCCTGTGAACC		
exon 18	TGTGGATTTGTGTGCTACGC	258	59
	CATAAATAGCACCGACAGGGA		
exon 19	GGGATGGAGAAGAGTGAGGA	159	57
	TCCTCACTCTTCTCCATCCC		
exons 20-21	ACCCTGTATGTTGCCTTGG	333	61
	GGGGATTTTGCTGTGTGCTG		
exon 22	ATTCCTGCTCCCACCGTCTC	282	57
	CACAGAGTGTCCGAGAGGCCA		
exons 22-23	GGAGATTATCAGGTGAGATGCC	608	61
	CAGAGTGTCCGAGAGGCAGGG		
exon 24	CGTTGACCCCTCCACCTTGA	375	58
	GGGAAAACATGCACCTTCTT		
polyadenylation signal	TAGGGGGTAAAATGGAGGAG	413	56
	ACTAACTCAGTGGAATAGGG		
	GGAGCTAGGATAGCTCAAT		

PCR products made on DNA from blood of specific LGMD2A patients were then subjected either to heteroduplex analysis or to direct sequencing, depending on whether the mutation, based on haplotype analysis, was expected to be homozygous or heterozygous, respectively. It was occasionally necessary to clone the PCR products to precisely identify the mutations (i.e., for microdeletions or insertions and for some heterozygotes). Disease-associated mutations are summarised in Table 4 hereunder and their position along the protein is shown in Fig. 4.

**Table 4:** nCL1 mutations in LGMD2A families.

Codons and amino acid positions are numbered on the basis of the cDNA sequence starting from ATG.

Exon	Families	Nucleotide position	Nucleotide change	Amino acid position	Amino acid change	Restriction si
2	B519*	328	<u>C</u> GA-> <u>T</u> GA	110	Arg->stop	
4	M42	545	<u>CT</u> G -> <u>C</u> AG	182	Leu->Gln	
4	M1394; M2888	550	CAA -> CA	184	frameshift	
5	M35; M37	701	GGG -> GAG	234	Gly->Glu	

15						
6	M32	945	CGG -> CG	315	frameshift	-SmaI
8	M2407*	1061	GIG -> GGG	354	Val-> Gly	
8	M1394	1079	TGG -> TAG	360	Trp->stop	-BstNI, -Eco
11	M2888	1468	CGG -> TGG	490	Arg->Trp	
13	R12*	1715	CGG -> CAG	572	Arg->Gln	-MspI
19	R27	2069-2070	deletion AC	690	frameshift	
21	R14; R17	2230	AGC -> GGC	744	Ser->Gly	-AluI
22	A*: B501*: M32	2306	CGG -> CAG	769	Arg->Gln	
22	B505	2313-2316	deletion AGAC	771-772	frameshift	
22	R14; B505	2362-2363	AG -> TCATCT	788	frameshift	

The first letter of the family code refers to the origin of the population B= Brazil, M= metropolitan France, R = Isle of La Réunion, A= Amish.

Each mutation was confirmed by heteroduplex analysis, by sequencing of both strands in several members of the family or by enzymatic digestion when the mutation resulted in the modification of a restriction site. Segregation analyses of the mutations, performed on DNAs from all available members of the families, confirmed that these sequence variations are on the parental chromosome carrying the LGMD2A mutation. To exclude the possibility that the missense substitutions might be polymorphisms, their presence was systematically tested in a control population: none of these mutations was seen among 120 control chromosomes from the CEPH reference families.

#### EXAMPLE 5 :

#### **Analysis of families genes, chromosome-15 ascertained families**

The initial screening for causative mutations was performed on families, each containing a LGMD gene located on chromosome 15. These included families from the Island of La Réunion (Beckmann et al., 1991), from the Old Order Amish from northern Indiana (Young et al., 1992,) and 2 Brazilian families (Passos Bueno et al., 1993).

#### **a) Reunion Island families**

Genealogical studies and geographic isolation of the families from the Isle of La Réunion were suggestive of a single founder effect. Genetic analyses are,

however, inconsistent with this hypothesis as the families present haplotype heterogeneity. At least, six different carrier chromosomes are encountered, (with affected individuals in several families being compound heterozygotes). Distinct mutations corresponding to four of these six haplotypes have been identified  
5 thus far.

In family R14, exons 13, 21 and 22 showed evidence for sequence variation upon heteroduplex analysis (Fig. 6). Sequencing of the associated PCR products revealed (i) a polymorphism in exon 13, (ii) a missense mutation (A->G) in exon  
10 21 transforming the Ser<sup>744</sup> residue to a glycine in the loop of the second EF-hand in domain IV of the protein (Figure 4), and (iii) a frameshift mutation in exon 22. The exon 21 mutation and the polymorphism in exon 13 form an haplotype which is also encountered in family R17. Subcloning of the PCR products was necessary to identify the exon 22 mutation. Sequencing of several clones  
15 revealed a replacement of AG by TCATCT (data not shown). This frameshift mutation causes premature termination at nucleotide 2400 where an in frame stop codon occurs (Figure. 4).

The affected individuals in family R12 are homozygous for all markers of the LGMD2A interval (Allamand, submitted). Sequencing of the PCR products of exon 13 revealed a G to A transition at base 1715 of the cDNA resulting in a  
20 substitution of glutamine for Arg<sup>572</sup> (Figure. 7) within domain III, a residue which is highly conserved throughout all known calpains. This mutation, detectable by loss of *MspI* restriction site, is present only in this family and in no other examined LGMD2A families or unrelated controls.

In family R27, heteroduplex analysis followed by sequencing of the PCR  
25 products of an affected child revealed a two base pair deletion in exon 19 (Figure. 6 and table 4). One AC out of three is missing at this position of the sequence, producing a stop codon at position 2069 of the cDNA sequence (Figure 4).

#### b) Amish families

30 As expected, due to multiple consanguineous links, the examined LGMD2A Northern Indiana Amish patients were homozygous for the haplotype on the chromosome bearing the mutant allele (Allamand, submitted). A (G->A) missense mutation was identified at nucleotide 2306 within exon 22 (Fig. 7). The

resulting codon change is CGG to CAG, transforming Arg<sup>769</sup> to glutamine. This residue, which is conserved throughout all members of the calpain family in all species, is located in domain IV of the protein within the 3rd EF-hand at the helix-loop junction (ref). This mutation was encountered in a homozygous state  
5 in all patients from 12 chromosome 15-linked Amish families, in agreement with the haplotype analysis. We also screened six Southern Indiana Amish LGMD families, for which the chromosome 15 locus was excluded by linkage analyses (Allamand ESHG, submitted, ASHG 94). As expected, this nucleotide change was not present in any of the patients from these families, thus confirming the  
10 genetic heterogeneity of this disease in this genetically related isolate.

c) Brazilian families

As a result of consanguineous marriages, two Brazilian families (B501, B519) are homozygous for extended LGMD2A carrier haplotypes (data not shown). Sequencing PCR products from affected individuals of these families  
15 demonstrated that family B501 has the same exon 22 mutation found in northern Indiana Amish patients (Figure 7), but embedded in a completely different haplotype. In family B519, the patients carry a C to T transition in exon 2, replacing Arg<sup>328</sup> with a TGA stop codon (Figure 7), thus leading, presumably, to a very truncated protein (Figure 4).

20 d) Analysis of other LGMD families

Having validated the role of the candidate gene in the chromosome 15 ascertained families, we next examined by heteroduplex analysis LGMD families for which linkage data were not informative. These included one Brazilian (B505) and 13 metropolitan French pedigrees.

25 Heteroduplex bands were revealed for exons 1, 3, 4, 5, 6, 8, 11, 22 of one or more patients (Figure 6). Of all sequence variants, 10 were identified as possible pathogenic mutations (5 missense, 1 nonsense and 4 frameshift mutations) and 3 as polymorphisms with no change of amino acid of the protein. All causative mutations identified are listed in Table 4 here-above. Identical  
30 mutations were uncovered in apparently unrelated families. The mutations shared by families M35 and M37, and M2888 and M1394, respectively, are likely to be the consequence of independent events since they are embedded in different marker haplotypes. In contrast, it is likely that the point mutation in exon

22 of the Amish and in the M32 kindreds corresponds to the same mutational event as both chromosomes share a common four marker haplotype (774G4A1-774G4A10-774G454D-774G4A2) around nCL1 (data not shown), possibly reflecting a common ancestor. The same holds true for the AG to TCATCT substitution mutation encountered in exon 22 in families B505 and R14. The exon 8 (T->G) transversion is present in the two carrier chromosomes of M2407, the only metropolitan family homozygous by haplotype, possibly reflecting an undocumented consanguinity. For some families, no disease-causing mutation has been detected thus far (M40 for example).

In addition to the polymorphism present in exon 13 in families R14 and R17 (position 668) and in the intragenic microsatellites, four additional neutral variations were detected: a (T->C) transition at position 96, abolishing a *DdeI* restriction site in exon 1 in M31; a (C->T) transition in exon 3 (position 495) in M40 and in M37 forming a haplotype with the exon 5 mutation (in the former family, this polymorphism does not cosegregate with the disease); a (T->C) transition in the paternally derived promotor in M42 at position -428, which was also evidenced in healthy controls; and a variable poly(G) in intron 22 close to the splice site in families R20, R11, R19, M35 and M37. The latter is also present in the members of the CEPH families, but is not useful as a genetic marker as the visualisation and interpretation of mononucleotide repeat alleles is difficult.

In total, sixteen independent mutational events representing fourteen different mutations were identified. All mutations cosegregate with the disease in LGMD2A families. The characterised morbid calpain alleles contain nucleotide changes which were not found in alleles from normal individual. The discovery of two nonsense and five frameshift mutations in nCL1 supports the hypothesis that a deficiency of this product causes LGMD2A. All seven mutations result in a premature in-frame stop codon, leading to the production of truncated and presumably inactive proteins (Figure 4). Evidences for the morbidity of the missense mutations come from (1) the relative high incidence of such mutations among LGMD2A patients ; although it is difficult in the absence of functional assays to differentiate between a polymorphism and a morbid mutation, the occurrence of different "missense" mutations in this gene cannot all be



accounted for as rare private polymorphisms; (2) the failure to observe these mutations in control chromosomes; and (3) the occurrence of mutations in evolutionarily conserved residues and/or in regions of documented functional importance. Four of seven missense mutations change an amino acid which is conserved in all known members of the calpain family in all species (Figure 3). Two of the remaining mutations affect less conserved amino acid residues, but are located in important functional domains. The substitution V354G in exon 8 is 4 residues before the asparagine at the active site and S744G in exon 21 is within the loop of the second EF-hand and may impair the calcium-dependent regulation of calpain activity or the interaction with a small subunit (Figure 4). Several missense mutations change a hydrophobic residue to a polar one, or vice versa (Table 4) possibly disrupting higher order structures.

## METHODS

### Description of the patients

The LGMD2A families analysed were from 4 different geographic origins. They included 3 Brazilian families, 13 interrelated nuclear families from the Isle of la Réunion, 10 French metropolitan families and 12 US Amish families. The majority of these families were previously ascertained to belong to the chromosome 15 group by linkage analysis (Beckmann, 1991; Young, Passos-Bueno et al., 1993). However, some families from metropolitan France as well as one Brazilian family, B505, had non significant lodscores for chromosome 15. Genomic DNA was obtained from peripheral blood lymphocytes.

### Sequencing of cosmid c774G4-1F11 and EcoRI restriction map of cosmids.

Cosmid 1F11 (Figure 1C) was subcloned following DNA preparation through Qiagen procedure (Qiagen Inc., USA) and partial digestion with either *Sau3A*, *RsaI* or *AclI*. Size-selected restriction fragments were recovered from low-melting agarose and eventually ligated with M13 or Bluescript (Stratagene, USA) vectors. After electroporation in *E.coli*, recombinant colonies were picked in 100 µl of LB/ampicillin media. PCR reactions were performed on 1 µl of the culture in 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.1% Triton X-100, 0.01 gelatine, 200µM of each dNTP, 1 U of Taq Polymerase (Amersham) with 100 ng of each vectors primers. Amplification was initiated by 5 min denaturation at 95°C, followed by 30 cycles of 40 sec denaturation at 92°C and 30 sec annealing

at 50°C. PCR products were purified through Microcon devices (Amicon, USA) and sequenced using the dideoxy chain termination method on an ABI sequencer (Applied Biosystems, Foster City, USA). The sequences were analysed and alignments performed using the XBAP software of the Staden package, version 93.9 (Staden, 1982). Gaps between sequence contigs were filled by walking with internal primers. *EcoRI* restriction map of cosmids was performed essentially as described in Sambrook et al. (1989).

#### Northern Blot analysis

The probes were labelled by random priming with dCTP-( $\alpha^{32}\text{P}$ ). Hybridisation was performed to human multiple tissue northern blots as recommended by the manufacturer (Clontech, USA).

#### Analysis of PCR products from LGMD2A families

One hundred ng of human DNA were used per PCR under the buffer and cycle conditions described in Fougerousse (1994) (annealing temperature shown in Table 3). Heteroduplex analysis (Keene et al., 1991) was performed by electrophoresis of ten  $\mu\text{l}$  of PCR products on a 1.5 mm-thick Hydrolink MDE gels (Bioprobe) at 500-600 volt for 12-15 h depending of the fragment length. Migration profile was visualised under UV after ethidium bromide staining.

For sequence analysis, the PCR products were subjected to dye-dideoxy sequencing, after purification through microcon devices (Amicon, USA). When necessary, depending on the nature of the mutations (e.g., frameshift mutation or for some heterozygotes), the PCR products were cloned using the TA cloning kit from Invitrogen (UK). One  $\mu\text{l}$  of product was ligated to 25 ng of vector at 12°C overnight. After electroporation into XL1-blue bacteria, several independent clones were analysed by PCR and sequenced as described above.

The invention results from the finding that the nCL1 gene when it is mutated is involved in the etiology of LGMD2A. It is exactly the contrary to what is stated in the litterature, e.g. that the disease is accompanied by the presence of a deregulated calpain. Identification of nCL1 as the defective gene in LGMD2A represents the first example of muscular dystrophy caused by mutation affecting a gene which is not a structural component of muscle tissue, in contrast with previously identified muscular dystrophies such as Duchenne and Becker (Bonilla et al., 1988), severe childhood autosomal recessive (Matsumara et al.,

1992), Fukuyama (Matsumara et al., 1993) and merosin-deficient congenital muscular dystrophies (Tomé et al., 1994).

The understanding of the LGMD2A phenotype needs to take into account the fact that there is no active nCL1 protein in several patients, a loss compatible with the recessive manifestation of this disease. Simple models in which this protease would be involved in the degradation or destabilisation of structural components of the cytoskeleton, extracellular matrix or dystrophin complex can therefore be ruled out. Furthermore, there are no signs of such alterations by immunocytogetic studies on LGMD2 muscle biopsies (Matsumara et al., 1993; Tomé et al., 1994). Likewise, since LGMD2A myofibers are apparently not different from other dystrophic ones, it seems unlikely that this calpain plays a role in myoblast fusion, as proposed for ubiquitous calpains (Wang et al., 1989).

All the data disclosed in these examples confirm that the nCL1 gene is a major gene involved in the disease when mutated.

The fact that morbidity results from the loss of an enzymatic activity raises hopes for novel pharmaco-therapeutic prospects. The availability of transgenic models will be an invaluable tool for these investigations.

The invention is also relative to the use of a nucleic acid or a sequence of nucleic acid of the invention, or to the use of a protein coded by the nucleic acid for the manufacturing of a drug in the prevention or treatment of LGMD2.

The finding that a defective calpain underlies the pathogenesis of LGMD2A may prove useful for the identification of the other loci involved in the LGMDs. Other forms of LGMD may indeed be caused by mutations in genes whose products are the CANP substrates or in genes involved in the regulation of nCL1 expression. Techniques such as the two-hybrid selection system (Fields et al., 1989) could lend themselves to the isolation of the natural protein substrate(s) of this calpain, and thus potentially help to identify other LGMD loci.

The invention also relates to the use of all or a part of the peptidic sequence of the enzyme, or of the enzyme, product of nCL1 gene, for the screening of the ligands of this enzyme, which might be also involved in the etiology and the morbidity of LGMD2

The ligands which might be involved are for example substrate(s), activators or inhibitors of the enzyme.

The nucleic acids of the invention might also be used in a screening method for the determination of the components which may act on the regulation of the gene expression.

A process of screening using either the enzyme or a host recombinant cell, containing the nCL1 gene and expressing the enzyme, is also a part of the invention.

The pharmacological methods, and the use of nucleic acid and peptidic sequences of the invention are very potent applications.

The methods used for such screenings of ligands or regulatory elements are those described for example for the screening of ligands using cloned receptors.

The identification of mutations in the nCL1 gene provides the means for direct prenatal or presymptomatic diagnosis and carrier detection in families in which both mutations have been identified. Gene-based accurate classification of LGMD2A families should prove useful for the differential diagnosis of this disorder.

The invention relates to a method of detection of a predisposition to LGMD2 in a family or a human being, such method comprising the steps of :

- selecting one or more exons or flanking sequences which are sensitive in said family;
- selecting the primers specific for the or these exons or their flanking sequences, a specific example being the PCR primers of Table 3, or an hybrid thereof,
- amplifying the nucleic acid sequence, the substrate for this amplification being the DNA of the human being to be checked for the predisposition, and
- comparing the amplified sequence to the corresponding sequence derived from Figure 2 or Figure 8.

Table 2 indicates the sequences of the introns-exons junctions, and primers comprising in their structure these junctions are also included in the invention.

All other primers suitable for such RNA or DNA amplification may be used in the method of the invention.

In the same way, any suitable amplification method : PCR (for Polymerase Chain Reaction ®) NASBA ® (for Nucleic acid Sequence Based Amplification), or others might be used.

The methods usually used in the detection of one site mutations, like ASO (Allele specific PCR), LCR, or ARMS (Amplification Refactory Mutation System) may be implemented with the specific primers of the invention.

5 The primers, such as described in Tables 1 and 3, or including junctions of Table 2, or more generally including the flanking sequences of one of the 24 exons are also a part of the invention.

The kit for the detection of a predisposition to LGMD2 by nucleic acid amplification is also in the scope of the invention, such a kit comprises a least PCR primers selected from the group of :

- 10 a) in those described in table 1  
b) in those described in table 3  
c) those including the introns-exons junctions of Table 2.  
d) derived from primers defined in a),b) or c).

15 The nucleic acid sequence of claim 1 to 3 might be inserted in a viral or a retroviral vector, said vector being able to transfect a packaging cell line.

The packaging transfected cell line, might be used as a drug for gene therapy of LGMD2.

The treatment of LGMD2 disease by gene therapy is implemented by a pharmaceutical composition containing a component selected from the group of :

- 20 a) a nucleic acid sequence according to claims 1 to 4,  
b) a cell line according to claim 24,  
c) an aminoacid sequence according to claims 5 to 9.

## REFERENCES

- 5 Arikawa, E., Hoffman, E. P., Kaido, M., Nonaka, I., Sugita, H. and Arahata, K. (1991). The frequency of patients with dystrophin abnormalities in a limb-girdle patient population. *Neurology* 41, 1491-1496.
- 10 Bashir, R., Strachan, T., Keers, S., Stephenson, A., Mahjneh, I., Marconi, G., Nashef, L. and Bushby, K. M. D. (1994). A gene for autosomal recessive limb-girdle muscular dystrophy maps to chromosome 2p. *Hum. Mol. Genet.* 3, 455-457.
- 15 Beckmann, J. S., Richard, I., Hillaire, D., Broux, O., Antignac, C., Bois, E., Cann, H., Cottingham, R. W., Jr., Feingold, N., Feingold, J., Kalil, J., Lathrop, G. M., Marcadet, A., Masset, M., Mignard, C., Passos-Bueno, M. R., Pellerain, N., Zatz, M., Dausset, J., Fardeau, M. and Cohen, D. (1991). A gene for limb-girdle muscular dystrophy maps to chromosome 15 by linkage. *C. R. Acad. Sci. Paris. III* 312, 141-148.
- 20 Birnstiel, M. L., Busslinger, M. and Sturb, K. (1985). Transcription termination and 3' processing: The end is in site! *Cell* 41, 349-359.
- 25 Blackwell, T. K. and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* 250, 1104-1110.
- Bonilla, E., Samitt, C. E., Miranda, A. F., Hays, A. P., Salviati, G., DiMauro, S., Kunkel, L. M., Hoffman, E. P. and Rowland, L. P. (1988). Duchenne muscular dystrophy: deficiency of dystrophin at the muscle cell surface. *Cell* 54, 447-452.
- 30 Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* 212, 563-578.

Bushby, K. M. D. (1994). Limb-girdle muscular dystrophy. In Diagnostic criteria for neuromuscular disorders. A. E. H. Emery, ed. (Baarn, The Netherlands: ENMC), pp 25-31.

- 5 Croall, D. E. and Demartino, G. N. (1991). Calcium-activated neutral protease (calpain) system: structure, function, and regulation. *Physiol. Rev.* 71, 813-847.

Dynan, W. S. and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* 35, 79-87.

10

Emery, A. E. H. (1991). Population frequencies of inherited neuromuscular diseases - a world survey. *Neuromuscular Disorders* 1, 19-29.

- 15 Emori, Y., Ohno, S., Tobita, M. and Suzuki, K. (1986). Gene structure of calcium-dependent protease retains the ancestral organization of the calcium-binding protein gene. *FEBS lett.* 194, 249-252.

Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* 340, 245-246.

20

- Fougerousse, F., Broux, O., Richard, I., Allamand V., Pereira de Souza, A., Bourg N., Brenguier L., Devaud C., Pasturaud P., Roudaut C., Chiannilkulchai N., Hillaire D., Bui H., Chumakov I., Weissenbach J., Cherif D., Cohen D. and J. S. Beckmann (1994). Mapping of a chromosome 15 region involved in Limb-Girdle Muscular Dystrophy. *Hum. Mol. Genet.* 3, 285-293.

25

Goll, D. E., Thompson, V. F., Taylor, R. G. and Zalewska, T. (1992). Is Calpain activity regulated by membranes and autolysis or by calcium and calpastatin? *BioEssays* 14, 549-556.

30

Gosset, L. A., Kelvin, D. J., Sternberg, E. A. and Olson, E. (1989). A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol. Cell. Biol.* 9, 5022-5033.

Hirai, S., Kawasaki, H., Yaniv, M. and Suzuki, K. (1991). Degradation of transcription factors, c-Jun and c-Fos, by calpain. *FEBS lett.* 1, 57-61.

- 5 Imajoh, S., Kawasaki, H. and Suzuki, K. (1986). Limited autolysis of calcium-activated neutral protease (CANP): reduction of the  $\text{Ca}^{2+}$  requirement is due to the NH2-terminal processing of the large subunit. *J. Biochem.* 100, 633-642.

- 10 Jackson, C. E. and Carey, J. H. (1961). Progressive muscular dystrophy: autosomal recessive type. *Pediatrics* 77-84.

Keen, J., Lester D., Inglehearn, C., Curtis, A. and Bhattacharya, S. (1991). Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. *Trends Genet.* 7, 5.

15

Kosak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12, 857-872.

- 20 Lovett, M., Kere, J. and Hinton, L. M. (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* 88, 9628-9632.

- 25 Matsumara, K., Tomé F. M. S., Collin H., Azibi K., Chaouch M., Kaplan J-K., Fardeau M. and Campbell K., P. (1992). Deficiency of the 50K dystrophin-associated glycoprotein in severe childhood autosomal recessive muscular dystrophy. *Nature* 359, 320-322.

- 30 Matsumura, K., Nonaka, I. and Campbell, K. P. (1993). Abnormal expression of dystrophin-associated proteins in Fukuyama-type congenital muscular dystrophy. *Lancet* 341, 521-522.



- Minty, A. and Kedes, L. (1986). Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. *Mol. Cell. Biol.* 6, 2125-2136.
- 5 Miyamoto, S., Maki, M., Schmitt, M. J., Hatanaka, M. and Verma, I. M. (1994). TNF- $\alpha$ - induced phosphorylation of IKB is a signal for its degradation but not dissociation from NF-KB. *Proc. Natl. Acad. Sci. USA* *in press*.
- Morton, N. E. and Chung, C. S. (1959). Formal genetics of muscular dystrophy.  
10 *Am. J. Hum. Genet.* 11, 360-379.
- Murachi, T. (1989). Intracellular regulatory system involving calpain and calpastatin. *Biochemistry Int.* 18, 263-294.
- 15 Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. and Suzuki, K. (1984). Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein? *Nature* 312, 566-570.
- Ohno, S., Minoshima, S., Kudoh, J., Fukuyama, R., Shimizu, Y., Ohmi-Imajoh, S.,  
20 Shimizu, N., Suzuki, K. (1989). Four genes for the calpain family locate on four different chromosomes. *Cytogen. Cell Genet.* 51, 1054.
- Passos-Bueno, M.-R., Richard, I., Vainzof, M., Fougrousse, F., Weissenbach, J., Broux, O., Cohen, D., Akiyama, J., Marie, S. K. N., Carvalho, A. A.,  
25 Guilherme, L., Kalil, J., Tsanaclis, A. M., Zatz, M. and Beckmann, J. S. (1993). Evidence of genetic heterogeneity in the autosomal recessive adult forms of limb-girdle muscular dystrophy following linkage analysis with 15q probes in Brazilian families. *J. Med. Genet.* 30, 385-387.
- 30 Richard, I., Broux, O., Chiannikulchai, N., Fougrousse, F., Allamand, V., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., Lorenzo, F., Sebastiani-Kabatchis, C., Schultz, R. A., Polymeropoulos, M. H., Gyapay, G.,

Auffray, C. and Beckmann, J. (1994). Regional localization of human chromosome 15 loci. *Genomics* 23, 619-627.

5 Sambrook, J., Fritsh, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold spring Harbor Laboratory Press, Cold spring Harbor, USA.

10 Shapiro, M. and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15, 7155-7174.

15 Sorimachi, H., Imajoh-Ohmi, S., Emori, Y., Kawasaki, H., Ohno, S., Minami, Y. and Suzuki K. (1989). Molecular cloning of a novel mammalian calcium-dependant protease distinct from both m- and mu- type. Specific expression of the mRNA in skeletal muscle. *J. Biol. Chem.* 264, 20106-20111.

20 Sorimachi, H., Ishiura, S. and Suzuki, K. (1993a). A novel tissue-specific calpain species expressed predominantly in the stomach comprises two alternative splicing products with and without  $\text{Ca}^{2+}$ -binding domain. *J. Biol. Chem.* 268, 19476-19482.

25 Sorimachi, H., Toyama-Sorimachi, N., Saido, T. C., Kawasaki, H., Sugita, H., Miyasaka, M., Arahata, K., Ishiura, S. and Suzuki, K. (1993b). Muscle-specific calpain, p94, is degraded by autolysis immediately after translation, resulting in disappearance from muscle. *J. Biol. Chem.* 268, 10593-10605.

Staden, R. (1982). An interactive graphic program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* 10, 2951-2961.

30 Suzuki, K. and Ohno, S. (1990). Calcium activated neutral protease. Structure-function relationship and functional implications. *Cell Struct. Funct.* 15, 1-6.

- Tagle, D. A., Swaroop, M., Lovett, M. and Collins, F. S. (1993). Magnetic bead capture of expressed sequences encoded within large genomic segments. *Nature* 361, 751-753.
- 5 Tomé, F. M. S., Evangelista T., Leclerc A., Sunada Y., Manole E., Estournet B., Barois A., Campbell K. P. and Fardeau M. (1994). Congenital muscular dystrophy with merosin deficiency. *C. R. Acad. Sci. Paris* 317, 351-357.
- 10 Uberbacher, E. C. and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* 88, 11261-11265.
- Walton, J. N. and Nattrass, F. J. (1954). On the classification, natural history and traitment of the myopathies. *Brain* 77, 169-231.
- 15 Wang, K. W., Villalobo, A. and Roufogalis, B. D. (1989). Calmodulin-binding proteins as calpain substrates. *Biochem. J.* 262, 693-706.
- 20 Young, K., Foroud, T., Williams, P., Jackson, C. E., Beckmann, J. S., Cohen, D., Conneally, P. M., Tischfield, J. and Hodes, M. E. (1992). Confirmation of linkage of limb-girdle muscular dystrophy, type-2, to chromosome 15. *Genomics* 13, 1370-1371.

**CLAIMS**

1. A nucleic acid sequence comprising :

1) the sequence represented in Figure 8; or

2) the sequence represented in Figure 2; or

5 3) a part of the sequence of Figure 2 with the proviso that it is able to code for a protein having a calcium dependant protease activity involved in a LGMD2 disease ; or

4) a sequence derived from a sequence defined in 1), 2) or 3) by substitution, deletion or addition of one or more nucleotides with the proviso that  
10 said sequence still codes for said protease.

2. A nucleic acid sequence that is complementary to a nucleic acid sequence according to claim 1.

3. A nucleic acid sequence comprising in its structure a nucleotidic sequence according to claim 1 or 2, under the control of regulatory elements,  
15 and involved in the expression of calpain activity in a LGMD2 disease.

4. A nucleic acid sequence encoding the aminoacid sequence represented in Figure 2.

5. An amino acid sequence which is coded by a nucleic acid sequence according to claims 1 to 4, characterized in that it is a calcium dependent  
20 protease enzyme belonging to the calpain family, involved in the etiology of LGMD2.

6. An aminoacid sequence according to claim 5 or 6, characterized in that either it contains the sequence such as represented in Figure 2, or the amino acid sequence of Figure 2 modified by deletion, insertion and/or replacement of  
25 one or more amino acids with the proviso that such aminoacid sequence has the calpain activity involved in LGMD2 disease.

7. An amino acid sequence according to claim 5 or 6, characterized in that LGMD2 is LGMD2A.

8. A host cell unable to express a calpain enzyme activity, characterized in  
30 that it is transformed or transfected with a nucleic acid sequence comprising all or part of the nucleic acid sequence according to any one of claims 1 to 4.

9. Use of a nucleic acid according to one of claims 1 to 4 or a host cell according to claim 8 in the manufacturing of a drug for the prevention or the treatment of an LGMD2 disease.

10. Use of an amino acid sequence according to claims 5 to 6 in the manufacturing of a drug for the prevention or the treatment of an LGMD2 disease.

11. Use according to claims 10 or 11, characterized in that LGMD2 is LGMD2A.

12. Use of an amino acid sequence according to claims 5 to 7 for the screening of the ligands of said amino acid sequence, said ligand being selected in a group consisting of substrate(s), co-factors or regulatory components.

13. Use of a nucleic acid sequence according to one of claims 1 to 4 in a screening method for the determination of the components which may act on the regulation of gene expression of calpain.

14. Use of an host cell according to claim 8 in a screening method for the determination of components active on the expression of the calpain.

15. A method for detecting of a predisposition to a LGMD2 disease in a family or a human being, such method comprising the steps of :

- selecting one or more exons or their flanking sequences of the gene,
- selecting primers specific for these exons, or their flanking sequences, or an hybrid thereof,
- amplifying the nucleic acid sequences with these primers, the substrate for this amplification being the DNA of a human being; and
- comparing the amplified sequence to the corresponding sequence derived from Figure 2 or Figure 8.

16. The method according to claim 15, characterized in that the primers are those selected from the group of :

- a) those described in Table 1;
- b) those described in Table 3; and
- c) those including the introns-exons junctions of Table 2;
- d) those derived from the primers in a), b), or c).

17. The method according to claim 15 or 16, characterized in that LGMD2 is LGMD2A.

18. A kit for the detection of a predisposition to LGMD2 by nucleic acid amplification characterized in that it comprises primers selected from the group of :

- a) those described in Table 1;
- 5      b) those described in Table 3; and
- c) those including the introns-exons junctions of Table 2;
- d) those derived from the primers in a), b) or c).

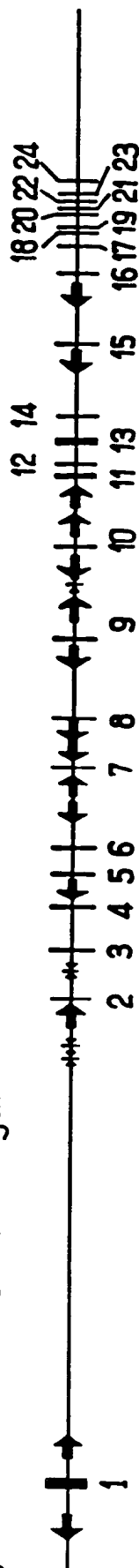
19. Use of a host cell according to claim 8 in a manufacturing of a drug for gene therapy of an LGMD2 disease.

10      20. Pharmaceutical composition for the treatment of an LGMD2 disease characterized in that it contains a component selected from the group of :

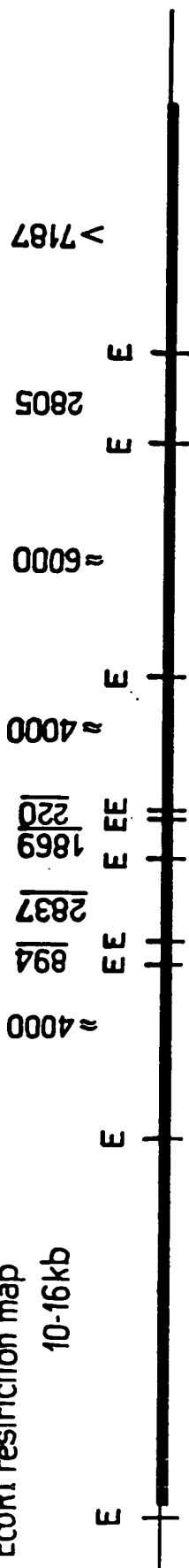
- a) a nucleic acid sequence according to claims 1 to 4,
- b) a host cell according to claim 8,
- c) an amino acid sequence according to claims 5 to 7.

**FIG.1**

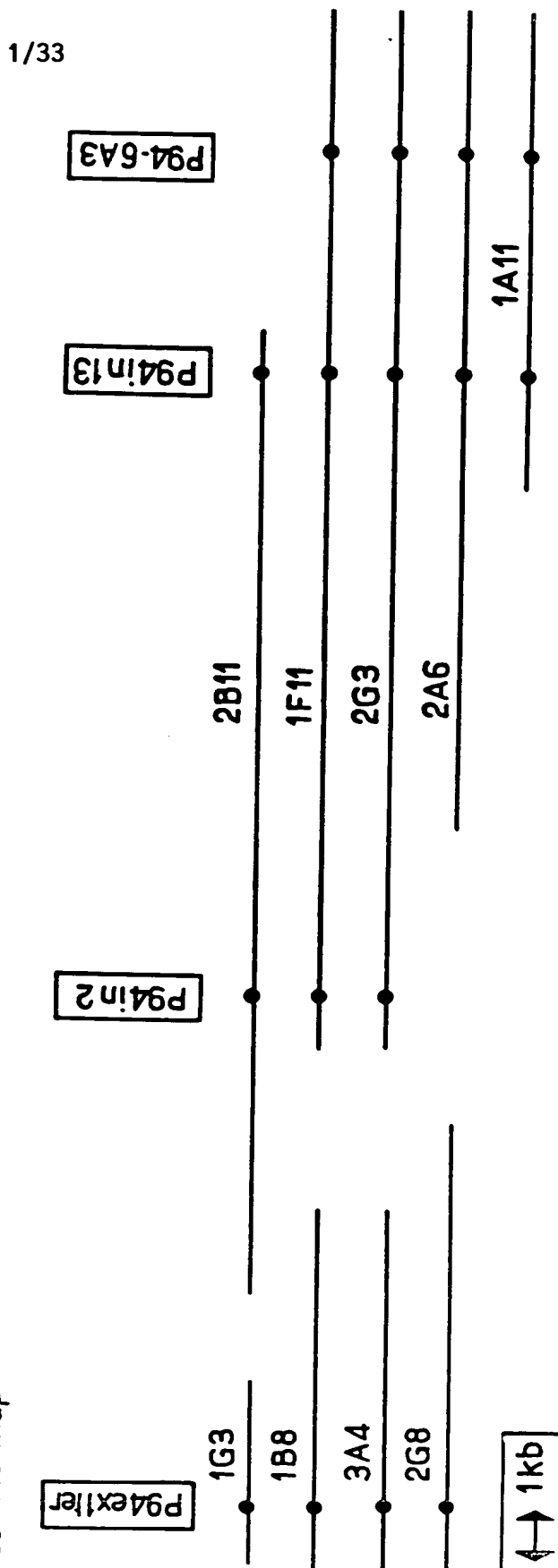
A) Genomic structure of the nCL1 gene



B) EcoRI restriction map  
10-16 kb



C) Cosmid map



**FIG. 2A**

[illegible]



**SUBSTITUTE SHEET (RULE 26)**

1  
ATGCCGACCGCTATTAGCGCATCTGTGGCTCCAAGCAGACGGCTGAGCCCGGTCCAGGGCCAGTTCTCCACCGGCCCAAGCCCACTGAGGCTGGGGTGGAAACCCAACT  
M P T V I S A S V A P R T A A E P R S P G P V P H P A Q S K A T E A G G G N P S 30 70 90 110

130  
GGCATCTATTCAGCCATCATACGCCGCAATTTCTCTATTATCGGAGTGAAGAGAGACATTGAGCACTTCACAGAAATGTCTAGAAAGAAAGTTCTTTATGTGGACCTGAGTTC  
G I Y S A I I S R N F P I I G V K E K T F E Q L H K K C L E K K V L Y V D P E F 150 170 190 210 230

250  
CCACCGATGAGACCTCTCTCTTTTATAGCCAGAGTTCCCATCCAGTTCGTCTGGAAGAGACCTCCGGAATTTCCGAGAAATCCCGCATTTATCATTTGATGGAGCCAAAGAACTGAC  
P P D E T S L F Y S Q K F P I Q F V W K R P P E I C E N P R F I I D G A N R T D 270 290 310 330 350

370  
ATCTGTCAAGCAGACTAGGAGCTGCTGGTTCTCGCAGCCATTGCTCGCTGACCTGAACGACCACTCTTTCCGAGTCATACCCCATGATCAAGTTTCATCGAAACTACGCA  
I C Q G E L G D C W F L A A I A C L T L N Q H L L F R V I P H D Q S F I E N Y A 410 430 450 470

490  
GGGATCTCCACTTCCACTCTCTGGCGTATGGAGAGTGGGTGGAGTGGTTATAGATGACTGCTGCTGCAAGCTTACAACTCAATCACTGGTTTTCACCAAGTCCACCAACCGCAATGAGTTC  
G I F H F Q F W R Y G E W V D V V I D D C L P T Y N N Q L V F T K S N H R N E F 510 530 550 570 590

610  
TCGAGTCTCTGCTGGAGAGGCTTATGCTAAGCTCCATGGTTCTACGAAGCTCTCAAGGTGGGAACACCAAGAGGCGCATGGAGGACTTCACAGAGGGGTGGCAGAGTTTTTGAG  
W S A L L E K A Y A K L H G S Y E A L K G G N T T E A H E D F T G G V A E F F E 630 650 670 690 710

730  
ATCAGGATGCTCTAGTGACATGTACAAGATCATGAAGAGCCATCGAGAGAGGCTCCCTCATGGCTGCTCCATTGATGGCAGCAATGACCTATGGAACCTCTCTCTCTGCT  
I R D A P S D H Y K I H K K A I E R G S L M G C S I D D G T N M T Y G T S P S G 750 770 790 810 830

850  
CTGAACATGGGGAGTTGATTGACCGCATGTGAAGGAATATGCAATACTCACTGCTCAGCACTGACACCTCCAGCCCAAGGCTCAGATCAAGAGCCGACCCGCAATCATTCGGTT  
L N H G E L I A R M V R N M D N S L L Q D S D L D P R G S D E R P T R T I I P V 870 890 910 930 950

970  
CAGTATGAGACAGAAATGGCGCTGCTCAGAGGTCAGCGCTACTGTCTGACGGGCTGGATGGTCCGTTCAAGGTGAGAAAGTGAAGCTGGTGGGCTGGGAAATCCGCTGG  
Q Y E T R M A C G L V R G H A Y S V T G L D E V P F K G E K V K L V R L R N P W 990 1010 1030 1050 1070

1090  
GGCAGGTGGAGTGAACGGTTCTTGGAGTCATATGGAAGCACTGAGCTTTGTGGACAAAGATCAGAGCCGCTCTCGACGACCAAGGCTCACTGAGATGGAGAGTCTGCAATGCC  
G Q V E W N G S W S D R W K D W S F V D K D E K A R L Q H Q V T E D G E F W M S 1110 1130 1150 1170 1190

1210  
TATCAGGATTTCAATTCACAAAGTTGGAGATCTGCAACCTCAGCGGCCATGCTCTGAGTCTGCAAGCTTCAGACCTGACAGTGTCTGTGAACGAGGCGCGCTGGGTACGG  
Y E D F I Y H F T K L E I C N L T A D A L Q S D K L Q T W T V S V N E G R W V R 1230 1250 1270 1290 1310

FIG. 2B/2

1330 ..... 1350 1370 1390 1410 1430  
 GGTTCCTGCGGAGGCTGCGGCACTCCCACTACTTCTGGACCAACCTCAGTACGCTGAAGCTCTGAGGAGGAGCAGTACCTGATGACTCGGAGGTGATTTGCAAGCTTC  
 C C S A G G C R N F P D T F W T N P Q Y R L K L L E E D D P D D S E V I C S F  
 1450 1470 1490 1510 1530 1550  
 CTGGTGGCCCTGATCCAGAAAGCGGCGGAGGAGCGGAGCTCTCTTCAACCATTTGGCTTGGCCATCTACGAGCTTCCCAAGAGATGCGACGGGAAACAGCGACACCTG  
 L V A L M Q K N R R K D R K L G A S L F T I G F A I Y E V P K E M H G N K Q H L  
 1570 1590 1610 1630 1650 1670  
 CAGAGCACTTCTTCTGTACAAAGCTCCAGGCAAGGCGGAGGAGCTACATCAACATGCGGAGGCTGCCAGGCTTCCGCTGCCAGGAGTACGTCATCGTGCCTCCAGC  
 Q K D F F L Y N A S K A R S K T Y I N M R E V S Q R F R L P P S E Y V I V P S T  
 1690 1710 1730 1750 1770 1790  
 TAGGAGCCCAAGGAGGCGGCAATTCATCTCCGGTCTTCTCTGMAAGAGGAACTCTCTGAGGAAGTTGAAATACCATCTCCGTGATCGGCCAGTGAAGAAAGAAACCAAG  
 Y E P H Q E G E F I L R V F S E K R N L S E E V E N T I S V D R P V K K K T K  
 1810 1830 1850 1870 1890 1910  
 CCCATCTTCTGTCGACAGCAAGCAAGCAAGGAGCTGGGTGGACCGAGGAGTCAAGAGGCGGCAAGGCAAGGCGCTGATNAGCAAGGAGTCCCCACAGCCACAG  
 P I I F V S D R A N S N K E L G V D Q E S E G K T S P D K Q K Q S P Q P Q  
 1930 1950 1970 1990 2010 2030  
 CCTGGAGCTCTGATCAGGAAGTGAGGACAGCAACAAATTCGGGAACATTTCAAGCAGATAGCAGGAGATCAGATGGAGATCTGTGCAGATGAGCTCAAGAGGTCTTTAACAAGTC  
 P G S S D Q E S E Q Q Q F R N I F K Q I A G D D H E I C A D E L K K V L N T V  
 2050 2070 2090 2110 2130 2150  
 GTCAACAAACAAAGGACCTGACAGACACAGGTTCACTGCGAGTCTGCGTACCATGATTCGCTCATGGATCAGATGGCTCTGGAAAGCTCAACCTGCGAGGAGTTCCACCACTG  
 V N K H K D L K T H G F T L E S C R S M I A L M D T D G S G K L N L Q E F H H L  
 2170 2190 2210 2230 2250 2270  
 TGGACAAAGATTAAAGCCTGGCAAAATTTTCAACACTATGACACAGACCAAGTCCGGACCATCAACAGTACGAGATCGGAAATCGCATCAGGACCGGAGTTCACCTCAACAG  
 W N K I K A W Q K I F K H Y D T D Q S G T I N S Y E M R N A V N D A G F H L N M  
 2290 2310 2330 2350 2370 2390  
 CAGCTCTATGACATCATTACCATGCGGTACGACACAAACATCAACATCGACTTTGACAGTTTCTGCTCTGTTAGGCTGGAGGCGATGTTCAAGCTTTTCATGCTTTGAC  
 Q L Y D I I I T M R Y A D K H M N I D F D S F I C C F V R L E G M F R A F H A F D  
 2410 2430 2450  
 AAGGATGGAGATGATATCAAGCTCAACGTTCTGGAATGGCTGACAGCTCACCATGTATGCTGA  
 K D G D G I I K L N V L E W L Q L T M Y A

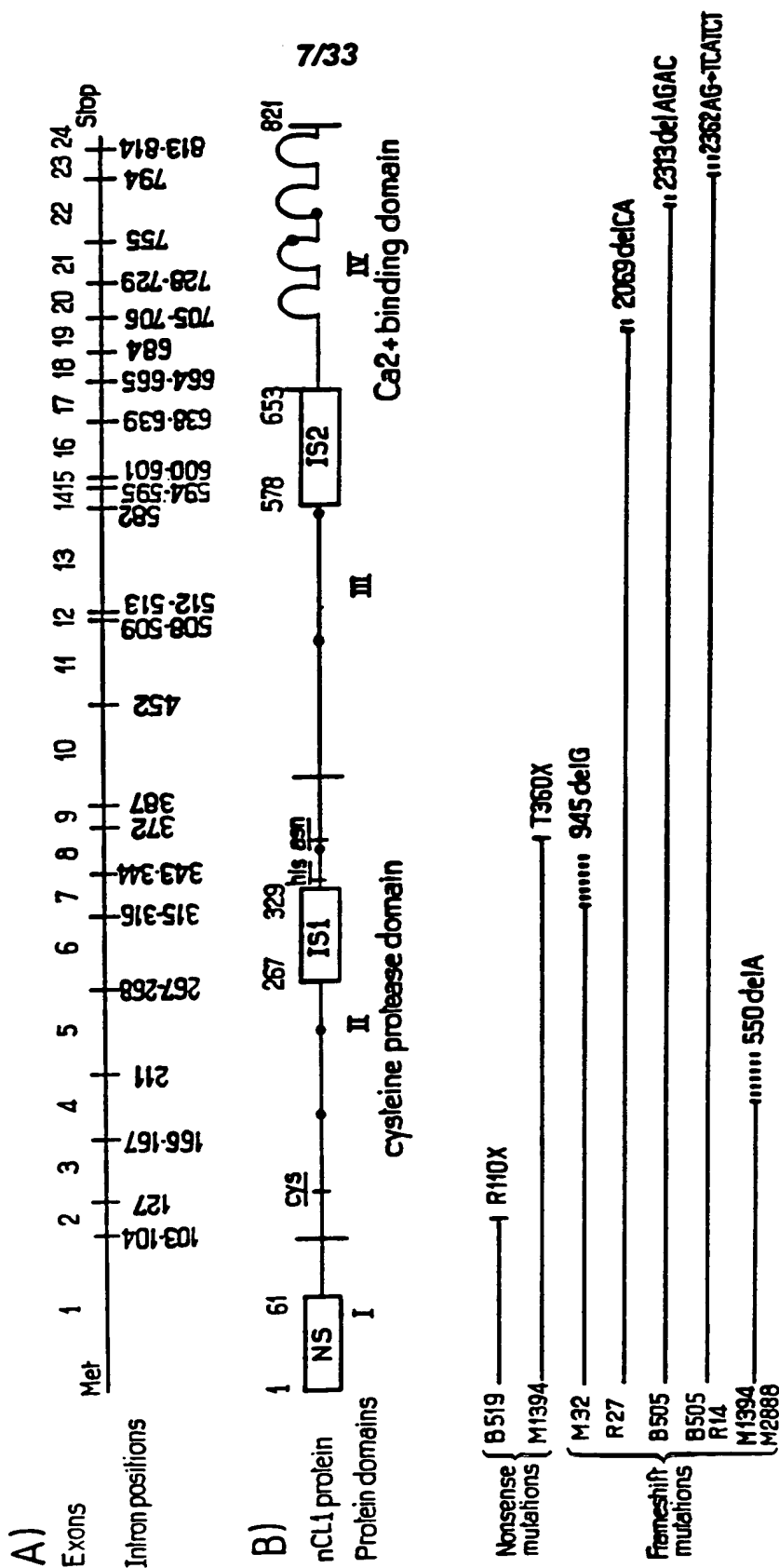
4/33



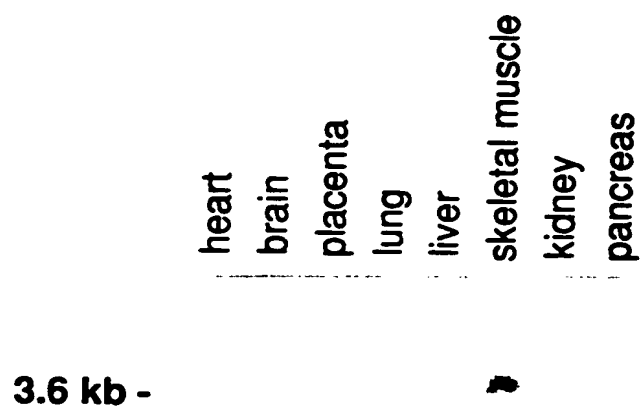
Figure 3:

Human	1	MPTVISASVAPRTAAERPSGVPVHPAOSKATEAGGNGPSGIYSALISRNPELIGVKEKTFE <sup>①</sup> HHKKCLEKKVLYVDPE <sup>②</sup> PPDETSFY <sup>③</sup> SQKFP <sup>④</sup> IQFV <sup>⑤</sup> WKREP	100
Rat	2	.....PT.....G.....G.T.....H.G.....L.....	
Pig	3		
Cow	4		
	1	EICENPFFI <sup>①</sup> IDG <sup>②</sup> ANR <sup>③</sup> PD <sup>④</sup> CGEL <sup>⑤</sup> GC <sup>⑥</sup> MF <sup>⑦</sup> LA <sup>⑧</sup> AIACLT <sup>⑨</sup> LN <sup>⑩</sup> OHLLFR <sup>⑪</sup> VP <sup>⑫</sup> HD <sup>⑬</sup> Q <sup>⑭</sup> SFIEN <sup>⑮</sup> YAGIF <sup>⑯</sup> HE <sup>⑰</sup> QF <sup>⑱</sup> WRY <sup>⑲</sup> GE <sup>⑳</sup> W <sup>㉑</sup> V <sup>㉒</sup> VD <sup>㉓</sup> CL <sup>㉔</sup> PT <sup>㉕</sup> YNNQ <sup>㉖</sup> VE <sup>㉗</sup> TK <sup>㉘</sup> SNHRNEEWSALL <sup>㉙</sup> E	200
	2	.....G.....D.....L.....ER.....T.....D.....	
	1	KAVAK <sup>①</sup> LHG <sup>②</sup> SY <sup>③</sup> EAL <sup>④</sup> KGGNT <sup>⑤</sup> EAM <sup>⑥</sup> ED <sup>⑦</sup> FG <sup>⑧</sup> VAE <sup>⑨</sup> FF <sup>⑩</sup> EIR <sup>⑪</sup> DA <sup>⑫</sup> ES <sup>⑬</sup> DMYKIMKKAIERGS <sup>⑭</sup> MG <sup>⑮</sup> CSID <sup>⑯</sup> DG <sup>⑰</sup> TN <sup>⑱</sup> TYGT <sup>㉑</sup> SPSGINMGELIARVVRNNDNSLLODS <sup>㉒</sup> LD <sup>㉓</sup> PRGS	300
	2	.....T.....K.....R.....	
	3		
	4		
	1	DERPRT <sup>①</sup> LIPVOX <sup>②</sup> ETRMACG <sup>③</sup> IVRG <sup>④</sup> HAYS <sup>⑤</sup> VG <sup>⑥</sup> LDEV <sup>⑦</sup> PF <sup>⑧</sup> GE <sup>⑨</sup> KV <sup>⑩</sup> TV <sup>⑪</sup> RLN <sup>⑫</sup> PRQ <sup>⑬</sup> VEW <sup>⑭</sup> NGS <sup>⑮</sup> SD <sup>⑯</sup> RWKD <sup>⑰</sup> WS <sup>⑱</sup> FVDKDEKARLQHQVTE <sup>㉑</sup> DGE <sup>㉒</sup> FW <sup>㉓</sup> MSY <sup>㉔</sup> ED <sup>㉕</sup> FIY <sup>㉖</sup> HFT <sup>㉗</sup> KL <sup>㉘</sup> E	400
	2	.D.S...V.....E.AL.....G.....	
	3	.D.....V...F.....E.AL.....S.....	
	4	.D.....M.V...F.....E.ALY.....S.....Y.....	
	1	IG <sup>①</sup> MTADAL <sup>②</sup> QSDK <sup>③</sup> LQ <sup>④</sup> TWTVSVNE <sup>⑤</sup> GR <sup>⑥</sup> VG <sup>⑦</sup> CCSAG <sup>⑧</sup> GCR <sup>⑨</sup> FP <sup>⑩</sup> DT <sup>⑪</sup> W <sup>⑫</sup> TNP <sup>⑬</sup> QV <sup>⑭</sup> RLK <sup>⑮</sup> LLEE <sup>⑯</sup> DD <sup>⑰</sup> PD <sup>⑱</sup> DD <sup>㉑</sup> SEVI <sup>㉒</sup> CS <sup>㉓</sup> FLVA <sup>㉔</sup> MA <sup>㉕</sup> KNRRKDRKI <sup>㉖</sup> CS <sup>㉗</sup> SLFT <sup>㉘</sup> IG <sup>㉙</sup> FAI <sup>㉚</sup> YEV <sup>㉛</sup> KE <sup>㉜</sup> MHG	500
	2	.....E.....TG.....	
	3	.....E.....TG.....	
	1	NKQ <sup>①</sup> QKDF <sup>②</sup> LYNASKARSKTYI <sup>③</sup> N <sup>④</sup> EV <sup>⑤</sup> SQ <sup>⑥</sup> RFL <sup>⑦</sup> PE <sup>⑧</sup> SE <sup>⑨</sup> VIV <sup>⑩</sup> PS <sup>⑪</sup> TYE <sup>⑫</sup> HQEGE <sup>⑬</sup> FI <sup>⑭</sup> RV <sup>⑮</sup> SEK <sup>⑯</sup> NLSE <sup>⑰</sup> VENTIS <sup>⑱</sup> VDR <sup>㉑</sup> PV <sup>㉒</sup> KKK <sup>㉓</sup> TK <sup>㉔</sup> PLIF <sup>㉕</sup> VS <sup>㉖</sup> DRANS <sup>㉗</sup> NKEL <sup>㉘</sup> GVD	600
	2	.....R.....E.....M.....K.....R.....	
	3	.....R.....E.....M.....K.....R.....	
	1	OESEEGKGTSPDKOKOSPOROPORGS <sup>①</sup> SDO <sup>②</sup> ES <sup>③</sup> EQOQ <sup>④</sup> FRNIFKQIAGDNE <sup>⑤</sup> CADE <sup>⑥</sup> KKV <sup>⑦</sup> NTV <sup>⑧</sup> VNKH <sup>⑨</sup> KDLK <sup>⑩</sup> TH <sup>⑪</sup> FT <sup>⑫</sup> LESCRS <sup>⑬</sup> MIALM <sup>⑭</sup> DT <sup>⑮</sup> DGSG <sup>⑯</sup> KLNL <sup>⑰</sup> Q <sup>⑱</sup> EE <sup>㉑</sup> HH <sup>㉒</sup> LM	700
	2	..A...D..G...GE...R..HT.....	
	3	.....QD.....EK..K.E.SNT.....	
	1	NKIKAWQK <sup>①</sup> FKKH <sup>②</sup> DT <sup>③</sup> DQSG <sup>④</sup> TINS <sup>⑤</sup> YEM <sup>⑥</sup> RV <sup>⑦</sup> AVND <sup>⑧</sup> AG <sup>⑨</sup> FHLN <sup>⑩</sup> NQ <sup>⑪</sup> LYDIIT <sup>⑫</sup> YAD <sup>⑬</sup> KHMNID <sup>⑭</sup> ED <sup>⑮</sup> SI <sup>⑯</sup> CC <sup>⑰</sup> FV <sup>⑱</sup> RE <sup>㉑</sup> EG <sup>㉒</sup> RA <sup>㉓</sup> FA <sup>㉔</sup> FD <sup>㉕</sup> KDGD <sup>㉖</sup> GI <sup>㉗</sup> IK <sup>㉘</sup> LN <sup>㉙</sup> VLEW <sup>㉚</sup> QLTMYA	800
	2	K.....H.....S.....	

FIG. 4

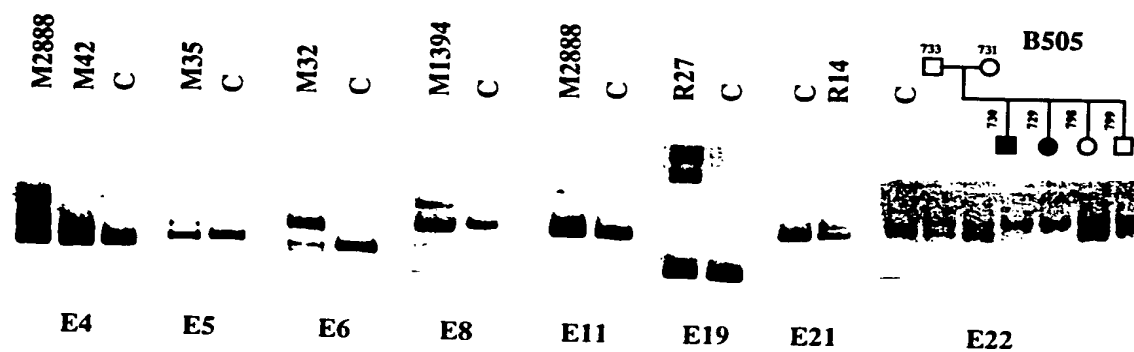


8/33



**FIG. 5**

9/33

**FIG. 6**

10/33

FIG. 7A) EXON 2Normal  
sequence

AATCCCCGATTTA

**B519**CGA → TGA  
Arg110 Stop

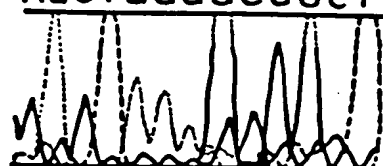
AATCCCT\*GATTTA

B) EXON 8Normal  
sequence

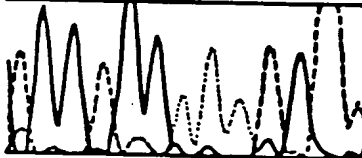
AGCTGGTGCGGCT

**M2407**GTG → GGG  
Val354 Gly

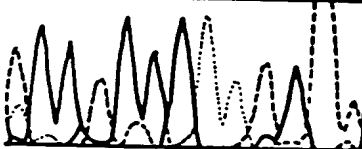
AGCTGGG\*GCGGCT

C) EXON 13Normal  
sequence

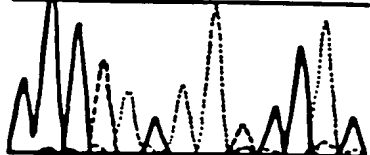
TCCTCCGGGTCTT

**R 12**CGG → CAG  
Arg572 Gln

TCCTCC\*AGGTCTT

D) EXON 22Normal  
sequence

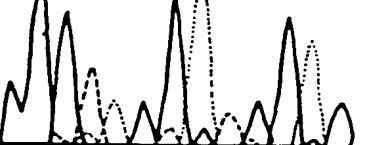
CCATGCGGTACGC

**Amish**CGG → CAG  
Arg769 Gln

CCATGC\*AGTACGC

**B 501**CGG → CAG  
Arg769 Gln

CCATGC\*AGTACGC





11/33

## LISTE DE SEQUENCES

## (1) INFORMATION GENERALE:

## (i) DEPOSANT:

- (A) NOM: AFM
- (B) RUE: 13, place de Rungis
- (C) VILLE: PARIS
- (E) PAYS: FRANCE
- (F) CODE POSTAL: 75013
- (G) TELEPHONE: (1) 45 65 13 00

(ii) TITRE DE L' INVENTION: LGMD GENE

(iii) NOMBRE DE SEQUENCES: 4

## (iv) FORME LISIBLE PAR ORDINATEUR:

- (A) TYPE DE SUPPORT: Floppy disk
- (B) ORDINATEUR: IBM PC compatible
- (C) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS
- (D) LOGICIEL: PatentIn Release #1.0, Version #1.25 (OEB)

## (2) INFORMATION POUR LA SEQ ID NO: 1:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 3018 paires de bases
- (B) TYPE: acide nucléique
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

TGATAGGTGC TTGTAACTG TGCTTAACGA AACATACCG TGTGCTGTAG GGA	60
CTTGTTTATA TCAGTTAGCC TGGTTTCGCT AACAGTACAT CATTTTGCTT AAAGTCACAG	120
CTTACGAGAA CCTATCGATG ATGTTAAGTG AGGATTTTCT CTGCTCAGGT GCACTTTTTT	180
TTTTTTTTTAA GACGGAGTCT CTTTCTGTCA CCTGGGCTGG AGTGCACTGG CGTGATCTGG	240
GTTCAACAACA ACCTCTGCCT CCTGGGTTCA AGCAATTCTT CTGTCTCAGC CTCCCAAGTA	300
GCTGGGATTA CAGGCACCCG CCGCCACACC CGGCTTATTT TTGTATTTTT AGTAGAGACA	360
GGGTTTCACT ATTGTTGACC ATGCTGGTCT CGAACTCGTG ACCTCATGTG ATCCACCCGC	420
CTCGGCCTCC CAAAGTGCAG AGATTAGAGA CGTGAGCCAC ATGGCCCAGC AGGACCACTT	480

FIG 8A/1

12/33

TTTAGCAGAT TCAGTCCCAG TGTTCAATTT GTGGATGGGG AGAGACAAGA GGTGCAAGGT	540
CAAGTGTGCA GGTAGAGACA GGGATTTTCT CAAATGAGGA CTCTGCTGAG TAGCATTTTC	600
CATGCAGACA TTTCCAATGA GCGCTGACCC AAGAACATT CAAAAAGATA CCAAATCTAA	660
CATTGAATAA TGTTCTGATA TCCTAAAATT TTAGGACTAA AAATCATGTT CTCTAAAATT	720
CACAGAATAT TTTTGTAGAA TTCAGTACCT CCCGTTACCC CTAAGTAGCT TTTTGTCAAT	780
ATTGTTTTCC ATTCATTGTA TGGGCAGTAG TTGGGTGGTC TGTATAACTG CCTACTCAAT	840
AACATGTCAG CAGTTCTCAG CTTCTTTCCA GTGTTACCT TACTCAGATA CTCCCTTTTC	900
ATTTTCTGTC AACACCAGCA CTTATGTCA ACAGAAATGT CCCTAGCCAG GTTCTCTCTC	960
TACCATGCAG TCTCTCTTGC TCTCATACTC ACAGTGTTC TTCACATCTA TTTTGTAGTT	1020
TCCTGGCTCA AGCATCTTCA GGCCACTGAA ACACAACCT CACTCTCTTT CTCTCTCCCT	1080
CTGGCATGCA TGCTGCTGGT AGGAGACCCC CAAGTCAACA TTGCTTCAGA AATCCTTTAG	1140
CACTCATTTT TCAGGAGAAC TTATGGCTTC AGAATCACAG CTCGGTTTTT AAGATGGACA	1200
TAACCTGTCC GACCTTCTGA TGGGCTTTCA ACTTTGAACT GGATGTGGAC ACTTTTCTCT	1260
CAGATGACAG AATTACTCCA ACTTCCCCTT TGCAGTTGCT TCCTTTCTTT GAAGGTAGCT	1320
GTATCTTATT TTCTTTAAAA AGCTTTTTCT TCCAAAGCCA CTTGCCATGC CGACCGTCAT	1380
TAGCGCATCT GTGGCTCCAA GGACAGCGGC TGAGCCCCGG TCCCAGGGC CAGTTCCTCA	1440
CCCGGCCCAG AGCAAGGCCA CTGAGGCTGG GGGTGGAAAC CCAAGTGGCA TCTATTCAGC	1500
CATCATCAGC CGCAATTTTC CTATTATCGG AGTGAAAGAG AAGACATTCT AGCAACTTCA	1560
CAAGAAATGT CTAGAAAAGA AAGTTCTTTA TGTGGACCCT GAGTTCCAC CGGATGAGAC	1620
CTCTCTCTTT TATAGCCAGA AGTTCCCAT CCAGTTCGTC TGCAAGAGAC TCCGGTGAGT	1680
AGTTCTCTGC TTGCTGGCTG GGTTCCCCC CCACGGAGGA GTCCTCTCAC TCAGCACCTC	1740
CGGCAGCTCA GCTGTGCACA TGGGCACTGG GGAAGGATC CTGGCAGCAG CTCTGCTGGG	1800
CTCTGTCTTT AAGTGTGAAG CAGGGAGGAG AGGAACAGGT CTCAGATATT TCACCAAATC	1860
TCAGCAAAAT CCAGAGGGAG AGCGCAGGAG GTGGGGTGAT TCTTATGCTC TGGCTCTTTC	1920
TCTCTGAAAA AAAAAAAAAA ATCTTGCTTT TTATAAAGT GGGTGGAAT CAGTTTAATT	1980
CATCCTGTAA AAATAAATAT TCCTTTCTCA GAACAAATTC CAGACAGCCC AGATGTACCT	2040
GTTCGTTTTA ATATTATTCA TCTTGTAAG ATTATTTTCT TTTCTCTGGC TAAAATCATG	2100

FIG 8A/2

SUBSTITUTE SHEET (RULE 26)

13/33

ATGTTATTCT TCTTTAATTT ACCAATGGCC ATTCTTTCTG AAACACAGAA ACCCTAGAAA	2160
GAGAAGAGTC ATAGGCAAGG AATTTTTTTC ATGCATAAAA TGTTGGGGTT AAAGAGAGAG	2220
AGACCTAGCA ATCGCTTTGG TCCACCTACC TCACCTCATA AGTGAGGAGT CAAGGCACAC	2280
TAGAGTGAAA TATATCTAGT GGGCACATGA CAGAGCCCCG ATTAAAACTT TGTTTTAGGA	2340
AACTCTCCCA GCCTCTGGGT TTCATTTACA GTGATCGCCA GGAGGGAAAT CACATTCCCC	2400
TGGCTCACCT CTCTGATCAT CCTCCAGTG TGACTCTTGT TCTTAATTCT AGAAATATTT	2460
ATTGAGCATC TACTAGTGCC AGCACTGGGC AAGCAACTGG GGGGACAGCA GTGAGTAAGA	2520
AAGACCAAAA TTCCAGCTGT CTTGGAACCT AGGGTCCTGA AGGGAAGATG GGCATTGAAC	2580
AAGAGTGACA TTGTCAGGAG ACGATGTTCT GGGTGCCACA GGATCATGTG GCAAGGAGAG	2640
CTAACCTGGT CCAGGGAGAC AAACCCTCTC TGAGGAAATG ATGACAAGCT GAGACCCAAT	2700
ACTATTGATT AGCCATGGTT TTCTTTAACC TAAGGTGGGC CAGGCATGGT GGCTCATGCC	2760
TATAAACCCA GCATTTTGA AGGCCAGGC TGGAGGATTG CTTGAGCCCA AGAGTTAGAG	2820
ACCAGCCTGG GCAACAGGCT GAAAACCTAT CTCTTTTGTA CTAAAAATTC AAAAAATTAT	2880
CCAGGCATGG TGGCACATGC CTGTGGTCCT AGCTACTCAG AGGCTGAGGT GGGAAGATCA	2940
CTTGAACCTG GGGAGTTTGA GGCAGCAGTG AGCCGAGATC ATGCCACTGC ACTCCAGGCT	3000
GGGTGACAGG AGTGAGAC	3018

14/33

## (2) INFORMATION POUR LA SEQ ID NO: 2:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 11451 paires de bases
- (B) TYPE: acide nucléique
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN (génomique)

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 2:

GATCCACCCG CCTTGGCCTC CCAAAGTGCT GAGATTACAG GTGTGAGCCA CCACGCCAC	60
CCGACACTGC CCTAACTCTC AAGTTGCATC CTTACTCGAA TAGTATGACA GTGTGGGAAG	120
CAGCATGGGA CAATGTAAAA AGGAGGCATG TTTCTGGCTT CTGCTACTTA CTAGCTGTGT	180
GTCTTTGCAC GAGTTTCTTA ACCTCTCTGG GCCTCAGTTT CCTTATCTGA AAAATAACAA	240
TGATAGTATT CCCTTCACAG GGCCAAATGG AATACTATCA GGAACACTAC ATAATGGAAC	300
TCAATAAATA ATAGCTACTG CGGCCGGGCG CGGTGGCTCA CATCTGTAAT CCCAGCACTT	360
TGGGAGGCCG AGGCGGGTGG ATCACAAGGT CAAGAGATGG AGACCATCCT GGCCAACATG	420
GTGAAACCGT ATCTCTACTA AAGATACAAA AATTAGCTGG GCATGGTGGC GCATGCCTAT	480
AGTCCCAGCT ACTCGAGAGG CTGAGGCAGG AGAATCACTT GAACCCCGGA GGCAGAGGTT	540
TCAGTGAGCC AAGATTGCAC CAGTGCCTG CAGCCTGGCG ACAGAGTGAG ACTCCGTCTC	600
AAAAAAATAC CTATCTATCT ATCTGTCTAT CTA CTGTTAT TCTTACCTGG TCATTTCTT	660
TTTGTTTCAC AGGAAATTTG CGAGAATCCC CGATTTATCA TTGATGGAGC CAACAGAACT	720
GACATCTGTC AAGGAGAGCT AGGTAGGAAA GTGCCTCAGG TCAGATCCTG CCAGATGATC	780
AAGGGGTGAT TACAAGGTGT GATCCCCTTC CAGGAGGTAA AGGGACAATC TGTGCTTGCT	840
TCCAGTAACT TTTTGGAAGA TTTTTTATAA CAGTTGCTTT ATGGTCGTTT ATCTACATGC	900
TGGCGATTGC TTCATTTCTT CCTACATGCC TCTTTAGCAC TCTGCCATGC ATCACAGGGG	960
GTATCTGCAT CCTGTGGCCT CCTCTCCAGT ATCTCAAGGA CACTTACATA CCCCCTCAG	1020
CATGACAAAA GCCCTGCTTT TCACTGTATC GTCTTTCTTG GAAGACAGCT CTGTGACTGT	1080
GCACCAAGCA TGCCCTTGG GCATGGAGAT TCTAGATACA CACACAAAAG GCATCGCCAA	1140
GGAAAGCACT TGTAAGTGA ACCCTTGTT TAAATTGGCC CAGCATAGCT CCATCTTTAA	1200

FIG. 8/B I

SUBSTITUTE SHEET (RULE 26)

15/33

AAGAGTCTTT CCACAAAGAT GGCATCCGCC ATGTGGATGA GCATCCAATT TTCTCTTTGA	1260
TTGGTTAGCT TGA CTGCTCC ATCTGATCTT CCTCTCTCTC GACCTCTTGT TCAGAAAGTA	1320
TTGTCTTTGG TGTGGACTAT AAGCAAGCTC TGTGAAGTAA AATTGGAGAG AACACCAACA	1380
GAAACAATTT AAATTTGAGG AAAAGGGGGC ACCTAAGACC AAAGGAATTT GGCTTATTTT	1440
ATTCCAGAAG GGGAGGCTGA GAATAAATCA GATGAATATC TGGGTTCTTG CACCTGAGGG	1500
AAGGCTTCCT GCAGAGCCCT GGGCATAATA ATCTGGGACC TTCAAACCAA TAACCTCTTT	1560
TCCAAGGAAA GACTGGCTGC TTCCAAGGAG GGTAGGGGAG AGTCGGGCTG CAGGCAGCTC	1620
TCAAGTCTCC CTTGACAC TCTCAGGTTG GCATTTTCAC TTTAACCCAT CCTCCCTTAA	1680
GAAGGCAGTT CTTTGTGACC AGGGTACACC CCCTATTATA TATATATATA CACACACAGA	1740
GAGAGAGAGA GAGAGAGAGA GAGAGAAAGA GAGCAAAGTG TTACCTCCAA CTACATACAG	1800
TACTCTGTCA GAAAAGAGGT TCAGAGAATA AGAAAACGTC CCGAGCTCAT TCCGTTGCCA	1860
GCAATGTCTT ACTGCCCCCT ATAGACGGGT TCCAGGGCAG CTGCCTACCT GGCCTTCCTT	1920
CCAATACAAA TCATCTTGGT GGATGGTTCT CTGAGGCTCA GTCTTCGCTG AAGTCAGAAG	1980
AGGAATTGGA CTCACATTGC AAAGGCACAG GGCAGGGCAG ATTTCTTACA GGTGTTAGGA	2040
AGAACAACCC AGTTATGATC ACCTACTGCT CTGTCTCCAT TGAGGCCTAA AAAGGAAGTG	2100
AGTTTATACT GCACTTGGAG GAACTGCCTG CAGCCTTGAG GAAAATGTCT AGTCACAAGG	2160
GAGTAAGTTA CCTGTTGATC ATATTGTCAA GGAATTCCTG TCCAATTCTC CTTCCCTGGG	2220
TTGACACCTC TGTAAGGTCA GATCTGGAAG TAGGAGAGTG GGCACCAAGG GAGTCCCCGT	2280
TCAGGGAAGT GGAGTGGCTG GCTGGGATTG GGGCTTTTTT TCCCAGGAG GAGCAGGAGT	2340
GCTCAGGATC TGTGCCCTGT GTCTGCCTGC AGGGGACTGC TGGTTTCTCG CAGCCATTGC	2400
CTGCCTGACC CTGAACCAGC ACCTTCTTTT CCGAGTCATA CCCCATGATC AAAGTTTCAT	2460
CGAAAACCTAC GCAGGGATCT TCCAATTCCA GGTGAGGTAA TGAGAGTGTA GTTAAGAGGG	2520
CCAGCGGCAG GCCACCCACC GCTGGTCTCC TGGCCTTGAC TTCCAGAAAG CTGGAGGAAA	2580
CTTCCCACCC ATCTACCCGC AGCGGCAACA GTCGGCATGG ACCCCCTTAA GGCTTCAAGC	2640
CTGGGAGGAA GCAGTTGCTT ATCTCTGGCT CCCTAATCCC TCCCCACCA CCTTCCACTA	2700
TGTCCCAGAA AGACAGGAAG ACATCCTGTT TACTGTGGGT CTATTTTTGT CTTTGCAGCT	2760
GTCTGGCTGC TTTTATTGCC TGCAGCCCTT CTCAAGTAGG TCCCTAAGAT ATTAGCACTG	2820

16/33

TGACACCACA GGACCCTTCA GGTGTACAG GAACCCCTGT CCAGGGCTCC TGTATACTTC	2880
TTCTCTCTA AGGCATGGCG GTACCAAGGC TATCACTCCT CTCTTCCAAG CCCTGGAAGA	2940
AGAGTCTGCT TAACCTGGGG ATCAGGCTTC TTGTTTGCCC TAGAACTGAA TCTGATGGTT	3000
CTAGAATCCA TCCAGCTACT GGAAATTTTC TGGGTCCCAG TCACCTTGGC ATAGAGCTGG	3060
TGCTAGAGCA GAACCAAACCT GAATTCTACC TGTGAGGGTC TCGTAGCTTC CGGGATGCTG	3120
GGGAGTCAGC CTGTCTCCAG CTTCAAAGGC TCCCTCATGT CCCAGGATGA CCCACATTAT	3180
CAGTTCTTGC TCCCCGGGTC TTGCACCTCA GCACGGAAGG CCTCAGAAAA GGTCTGTCTC	3240
CAGGCTCAGA CTCCCCCTCC TGCCGCCTTG GGAACATGGC ATATTAAAG GGTCTCAGAT	3300
CTAAAGGGCC TTACATACAA ATATCAGATA GATTTCTGTT CTCATTTCAG TGAGGGAGAA	3360
AGTGCCATTG AAAAGGAGAC TAAACCACAT TTGGCCCTTT TCAGTTCAAA CTGATTCAAT	3420
CAAAAAAGAG CGACATCCAA ACTTGAAATG ATTGAACAAT GTTCCTGCTA CAGCTAGAAT	3480
AGATTCTGGG TCACTTTGTT CCTCCGTTTC AATCCTTGTT CTTCAAGTTG GCATCAAGAA	3540
ATACCTAAAT CAGCACAGTG CCTTCACTGC ATAGTTCCCA ATCCTGGCCA CATTGAATCA	3600
GCTGGGGGCA CCTGAGAGTG CTGACACCCA GGCCCTGCCC CAGACCTGCT GAGCAGGAGA	3660
ATGAAAATCT TACATCCTAA GACACTCATG GAGCACCTAC TCTACCCATT ACTGGGCTGG	3720
ACTCTGTGGA AGACATGAAG TATATGTAAC TCACTTCCAG CTCTCAAAAA GCACCCAGTC	3780
CAGTTAGAGA CAGATTTACA CACCCCAAAC ACAAATAGG ATGAACAGGC ACCCAGATGC	3840
AGAGTCCAGG AAATGATGCT GCTTTGGGAT TCAAGAACCC CCTGAGGAAT GTGGAGGAAG	3900
GACACATTTT CTAACAGTAA TTTGAGTATG TGAATCTGTG CGTGACGCTT CTGTGCAGTT	3960
CTGGCGCTAT GGAGAGTGGG TGGACGTGGT TATAGATGAC TGCCTGCCAA CGTACAACAA	4020
TCAACTGGTT TTCACCAAGT CCAACCACCG CAATGAGTTC TGGAGTGCTC TGCTGGAGAA	4080
GGCTTATGCT AAGTAAGCAA CACTTTAGAA TGTGAGGTGG GCCTAGAGGT GAGAAAGTGG	4140
GTTGCAAAAT CCAGCCGAGA CCTCACTCAC AGGAAGAGGC ATGTGCCTCT ATACGTGCAT	4200
ATGTGTGGGC ATGCAAGTCC AACTGTGACC CAAAGTTAGA GATCAGTTCC AGGCAACAAC	4260
AGCTCTAACT AAAAACATTA AATTTAAGAG TAGAAATGAA GATTTGCATA GAAGACCTTT	4320
AGCTTTAGCT CACCATAGCG AGTTCTTTCA TTGCACCTCC ATGGTGGCAT TGCAAGTCTT	4380
GGGATCAGAG CATTGTCCCA GGGTCTCGAT TGGCTCAACC TCATGTGCTT ATAGAAGATT	4440

FIG. 8B/3

SUBSTITUTE SHEET (RULE 26)

17/33

TATAAAGACA TGTTGTCTCT CAACTTAAAA GCTCCACCCC AGATGATAAT AATGGATTTT	4500
CAAATTTTGG AACAAAGGTCA CTCTGTAATG CAGGCTGGAG TGCAGTGGTG CAGTCACGGA	4560
TCACTGTAGA TTGACCTCCT GGGTTCAAGG TGCTCCTCCC ACCTCAGCCT CCCAAGTAGC	4620
TGGGACTACA TGCGGGCATC ACCATGGCCC TTTTATTTTT GTATTTTTTT GTAGAGCGGG	4680
GTTTTCCCAT GTTGACCCAG ACTGTTCTCG AACTCTTGGG CTCATACAAT CCACCAGCCT	4740
TGCCCTCCCG AAGCGCTGGG ATTGCCGGTG TGAGCCACCA CACCGGCAGC TGCTAATGGC	4800
TTAATGCAG CCCTTCCTCA ACGTTCAGGA TGTAGTGGA AGAGCTCTCA GGAAGTGGGG	4860
ATAGCTGGGT TTCAATCCCA GTGCTTCTGG CTCTCTGTGG TCTTGGGTGG GTCACTTAGC	4920
CTCTTGAGCT CAGTTTCTTC ATTATGAAGA AAGGGAATCA TTGTTTCCAT CCCATGAGCT	4980
CATAGGGTTA ATGTGGAATT GATGAAAGAA CATCACAGCA TCCAAGAGGT AAAGTTCTGG	5040
TGGCAGTGGT ACCTGGGTTT TGTTCCCTGG AACTCTGTGA CCCCAAATTG GTCTTCATCC	5100
TCTCTCTAAG GCTCCATGGT TCCTACGAAG CTCTGAAAGG TGGGAACACC ACAGAGGCCA	5160
TGGAGGACTT CACAGGAGGG GTGGCAGAGT TTTTGAGAT CAGGGATGCT CCTAGTGACA	5220
TGTACAAGAT CATGAAGAAA GCCATCGAGA GAGGCTCCCT CATGGGCTGC TCCATTGATG	5280
TAAGTCTGGG GTGTGGGGCA CAGGGTGGGG AGCTCCAAGT GTCAGGAAGC CTTTTACCCA	5340
ATGAAGGGCA GCATAGAGCT TTTGTGTGGG ACAGAGCGAA TGTTTTGTTT GAGGAAGCAG	5400
GAAGTGGCTC TCAACTTTGA GGAAGTGGGA TTTCTCAAGG GAGAACAGTT CTTCCGGATT	5460
TTCAATAAAG AACTGGTCA AGGACATTTT AAGCCCTGGA ATGTCAGTGG AAATCAGTCC	5520
AGAGGCCTGT GTCAGTGGAG GCCTCCCTTG CTGGTGCTCC TCAGTCTCAG CACGCTCCCA	5580
TTAAGCTGGC CACGTACTTG GCTGTGGACC TGAGCCCACC ATTCCTTAA GAAAGCCTCC	5640
CAGTCACTGG GCTTTCACCA CACCTCCCCG CTTGAGACGT GGGCTTTGTG TTGTTACCTG	5700
GGAGAAGCTA AGCCTGCAGC ACCTTTCAGT GCAAAGAAAT GCTGTGAACT GAGACAGGAG	5760
CCAAGGGTAG GGAGATGGCC GCCCATGGCC AGGCCTCCTT CAGGGGGCAT GCCTTCCCTG	5820
AGGGCTGCTC AGTATATTGA TATGATAATC TTAGTGTTTT CCATTGGGGA GGATGGGGCT	5880
GAACTGAAT TCCTGCCCCT TCTTCTCCCA ACACGCCCAA TGGACAGCTT GGAAGGTCAG	5940
TTAGCACACA ACACCATGGA TGAACTTTTT TTCTGTATCA CTTTTCTCCG TCTTCTCTCC	6000
ATTCGTGCTC TGTTGATCTC TCCTCTCTCC CTTGTCTGT CCCATCTCTT TCTCTCTCT	6060

FIG. 8B/4

SUBSTITUTE SHEET (RULE 26)

18/33

CCTCCCTTT CCACCCTTCT GTGTTTGTT TCTCCCTCCC CTGTGTTGTT CCCTACATTC	6120
TCCATCGGGC CTCAGGATGG CACGAACATG ACCTATGGAA CCTCTCCTTC TGGTCTGAAC	6180
ATGGGGGAGT TGATTGCACG GATGGTAAGG AATATGGATA ACTCACTGCT CCAGGACTCA	6240
GACCTCGACC CCAGAGGCTC AGATGAAAGA CCGACCCGGG TGTGTACACC TCCGATTATC	6300
AGAACTGACC ATCCCTCCAA CCCACATGAC CCCGCCCTAT TAGTGTGAGA CTCCCCTCAG	6360
CAGCCAGGGC CTTACCCACA CACCCCCACC TGGCACCTCC CAAGGGTCTG GGTGAAATA	6420
ACTTGCTCAG CCAAGGCTCC TGAAGAGGGT GCAAGAACCA GGATTTTGGA GGAATCTCT	6480
GCTGGAGTTT CTGCATATTC CATGGTCCAG GCAGTTCCTC TCATAACGAA CTATCAGACA	6540
GAAATACTTG TAAAGATACT TCATTTATTT TGAAATATTT TTCCTCTTCT AATGTATTCA	6600
TTTATTCATT CAACACTTAT TTTGAGCTC CTAATATGTT CCAGGCACTC CTCTAGCAAA	6660
CAAAGCAAAT TCTCTCCTCT TTTTCAATAT TTGTGAAAA AGCAAGGTCT CCCTCTTGTA	6720
GAGTTTATAT TCTAGTATTT TCATAAGTTA TACCTGCTCA CTGGAGAATA CTGAGCCATA	6780
CAGAAAAACA CAGAGGAAAA TTTCACTTAT ATTTTCCCC ATGTAAAGAT AACCCTCTT	6840
AACATCTAGT ATATGTTCTT CCAGGATTTT TCTATGCACA CACTGAATCT GTATTTTTAT	6900
TTTTAAATG TTATCATATT GTATGTACCT CTTGTCAGCC TGCTTTTTTC AGTTAGTTTT	6960
TTTGTTTTT TGGTTTTTTT TTTTTTTTGG AAACCAAGTC TTGCTCTATT CCCTAGGCTG	7020
GAGCACAGTT GTTGCCATCT CGGCTCACTG CAACCTCTGC CTCCAAAGTT AAATAATTC	7080
TCCTGCCTCA GCCTCCCGAC ATAGCTGGGA TTACAGGCAC ACACCACCAC ACATGGCTAA	7140
TTTTTGTTAT TTTTAGTAGA GACGGGGTTT CACCATGTTG GCTGGAATGG TCTTGAAGTC	7200
CTGACCTCAA GTGATCCACC TGCCTCAGCC TCCCAAAGTG CTGGGATTAC AAGTGTAAGC	7260
CACCACACCC GGCCTAGTTT GATATTCTTA ATGTGCCCAA AGTATTCTCC TGTAACATTT	7320
TTTAATAGCT ACACAATATT CAAACACACA GATATGTTAT AATTTATTTA CCAATACCC	7380
TATTATTGGA AAGTTGAGTT CTTTTTTTTT TTTGTTTTGT TTTGTTTTGC TACTATTCTA	7440
AAATGCTATA ACGAACATCC CAATAGATAC ATCTTTGTAT ACATCCATGG TGAATCCAT	7500
AGGACAGATT CCCAGCAGTA GAATTGCTGG GTTGAATGAT ATGCTTAGGG TAATGACAGA	7560
AGAGTCATTT CAAGCAGCTT CCTAGGGTCT TAGAACTTAA GGATTAATGA GTCTTCCCGC	7620
CCCCTCCAG TCTATTCAGC ATGATCTGGA TCATGAGGAC TGAGATCTGG AAGAGACTGA	7680

FIG. 8B/5

SUBSTITUTE SHEET (RULE 26)



19/33

GATCTGGGAG AGGCTGAGAT ACCAAAAGCC CTGGCTCCAC CCATACCCCT CGCCCTGAAA	7740
ACAGCTCTAG GAATTCCGCG GCCTAGCAAG GCTCCGGGAA GCTCCTTTTA AAGCTGTGAC	7800
GTTAGTAGGC ACATGGACCA TAGAGACCTA TCCAGGGCTC ATGGGACTTT AGTGATCCTG	7860
CCCTTCTCCC AAGGATCCCC CATGGCTGCA ACTTGAAAT TTCTGCAAAT GGAAGAGCTA	7920
CTCCTTAGGC ACGGTCATGT CTGAGCAGGG ATCTCCTCGG GCTTTCTTAG AATTCTCTCC	7980
CTGGGCACTG GCACTCTTGA TTTCTTGAAT ATTATGTTCC AGGTGGGTGT GGAGGAGGTG	8040
AGGGGATGTA AAGAAGGCTA GACTTGGCCA GGCGCAGTGG CTCATGCCTG TAATCCCAGC	8100
ACTTTGGGAG GCTGAGGCGG GTGGATCACC TGAGGTCAGG AGTTCGAGAC CAGCCTGGCT	8160
AACATGGTGA AACCCCGTTT CTAATAAAAA TACAAAAAAT TAGCTGAGCA TGGTGGCAGC	8220
TGCCTGTAAT CCCAGCTACT CGGGAGGCTG AGGCAGGAGT ATCGCTGGAA CACGGGAGGC	8280
AGAGATTGCA GTGACCCGAG ATCGCGCCAC TGCACTCCAG CCTGGGCGAC ACAGCAAGAC	8340
TCTGTCTCAA AAAACAAAAA AGAAAGAAAA AAAGGAAAAG CTAAGACTTA CATGTGTCAC	8400
TTAACCCCTT TTCTCAAACC TCTTTCTCTT CCAGGAATAG TCAACCCCTG GATGGCTTCA	8460
GGGGAAGGGG GATCCTGAAG CCCAGGGCAG CCTCCAAC TC TACCCCTTCC TCCTTTGAAG	8520
GATACTAAGG GGTCCAGAAA GGAGGGGCAG GACACTGTTA CCCACCCAC ATCCAGCAT	8580
CCACATTGCT CTCTGATGGT CAGGACAGAG CCTTCTCAGG GAGACCAGCC TGTCTGGAGC	8640
TGTGTCTCTT GCACTCTTA AAGGGCCACT GAAGGTCCGT TCGTGGTCGT GAGGCACACT	8700
TTCAGGGAGC AGAGTGGTCT GTGTCTTCAC AGAGCCCGGA AAATGAACTA GTATGAACTT	8760
TGCCTCCAAG CAGCAGAACT TCTGTTCCCC CGCCCCTAAT GGGTTCTCTG GTTACTGCTC	8820
TACAGACAAT CATTCCGGTT CAGTATGAGA CAAGAATGGC CTGCGGGCTG GTCAGAGGTC	8880
ACGCCTACTC TGTCACGGGG CTGGATGAGG TAAGCCTGGT GGGGCTTGGT GGGGCAAGGG	8940
CACCCTCCTG GGTAAACCTC ATGAAGTCAG GACTTAGCTG TTGGGGCCCC TGCCCTGTCT	9000
GCAGAGCTTG CCTCCAATCA GGACATTCAG TTCAAGGTCC AAGCCACGCC TGGGAGCAGA	9060
GGGGCCTGTG AAAGTGGTAG AGGTGGATCC TGCCACAGTT GGTGCACAGT TTATCTTTGC	9120
TTTTCGTGCT AAAGATGGCA ATTTTTCCAA CATTTCCAAT GAACAAATTG AAATATCACT	9180
TAACTTTGCT TTTACAAAGT TGGTTTCATG TGTTCTTGAG CTTCTGTTC TCTCGTGTC	9240
AGATAGCTAC AGTTGTCTCT GGGTAGCCAC GGGGACTGGT TCCAGAAGCC CCAACAGTAA	9300

FIG. 8B/6

SUBSTITUTE SHEET (RULE 26)

20/33

CAAAATCTGC AGATGCTCAA GTCCCTTCTG TAAAATGGAG TAGTATTTGC ATATAACCTA	9360
TGCACATCCT CCCATATACT TTAAGTCATC TCTGGATTAC TTACGATACC TAACACAATG	9420
GAAATGCTAT GTAAATAGTT ATTGCACTGC ATTGGGTTTT TTTGGTATTA TTTTCTGTTG	9480
TTGTATTATT ATTTTTTCTT TTTTTGAATA TTTTGTATCC ACAATTGGTT ATATGCCAAA	9540
GCCATGGATA CGAGAGGCTG ACTGTTCTGT TTTGCTCCTT CTGGGACTTC TGGGTTTTCC	9600
TGGACCATGT CTGAGACAGG AACGTTGTAA GACCTGTTGC ACACAGTTGG GCAGGTTGTG	9660
CCCTGTACAG AGGGATGGGC TGAGAGGGGC AGTTGCCTGC ATCACCATT GCAGCAGACT	9720
GGAGGGAGTC TGCTTGTTTG TAGTTCCTCA GTCAGCAGGG GCCTTTTGTG TTTCTTCCT	9780
TTCTTTTTTT TTTTTTTTG AGACGGAGTC TCACTCTGTT GCCCAGGCTG GAGTGTAGTG	9840
GCACAGTCTC GGCTCACTGC AATGTCCGCC TCCTGGATTG AAGCGATTTT CCTGCCTCAG	9900
CCTCCTGAGT AGCTGGGATT ACAGGCGCGT GTCACCATGC CCAGCTAATT TTTGTATTTT	9960
TAGTAGAGAT GGGGGTTTCT CCATGTTGAT CAGGCTGGTC TCGAACTCCT GACCTCGTGA	10020
TCCGCCCACC TCGGCCTCTC AAAGTGCTGG GATTACAGGC GTGAGCCACC ACGCCTGGCC	10080
AGCAGGGGCC TTTTTTCTAA TTTATATGAA GACACCTAAT TTATATGTGT TAGCAAAGCC	10140
CTCCTGTTTA TGCCTCACCT CCTCCCCGA AGCTCATACG GCAGGATGTT CCTGAGAAAA	10200
TTGCCTCTTA GAAGATAGAG AGGAGATGCC AAGCCTAAGT TAGGCAGACT CAGGAGGATA	10260
GGTCTGACCC ACCCCTGCC ATTCCCAGC ACACTTGTA TTAATCTCCT TGGCCAGAGC	10320
CAGGCAGAAC ACCCTCGCGT AAGAGATTTG CCCCCAGCC CCGTCCCAGC CCTCAGCTAG	10380
ACAGAAGATT CCCTTTCAG AGAGGCTGCA GAGCATGAGA GCTCTTTCTG TGTGCTTAAG	10440
GTCCCGTTCA AAGGTGAGAA AGTGAAGCTG GTGCGGCTGC GGAATCCGTG GGGCCAGGTG	10500
GAGTGAACG GTTCTTGGAG TGATAGGTAG GTGAGGGGAC CCCACGGGAT TGGCGGTGGC	10560
GGGGAACAGG GTCCGGGACA AGGCTGTGTT GGGAACTGAG CCATGAGAGT ATTGAAGATG	10620
CTTGGTATAA AATCACCTC AAAACCAATG ATCCGCAGAG AAGAGGGGCA CAGGTGTTGG	10680
CTCCAGGGAA GGGCCAGGAG TGGAAGCGGG GTGCTGGGGA CCCAGAGAGG TTGCTGACAA	10740
CCATTGGCTG GAAAGGAAGG ATTCCAGAAA GCGTGGGGAA GGTCCAGGCA GGAAAAGCGT	10800
ATGAATGCAG GGTTCTGGGC TAGAGAAGTG ACTTCCCTTC TTGGGTCTT GTGTTGCCTT	10860
TCCTGTGAAA TGGGAACAGT ATTATTAGCA CTTACCTTGT GGGCTGATAT TGAGGAGTAA	10920

FIG.8B/7

SUBSTITUTE SHEET (RULE 26)

21/33

CTGGGACTTG TTTTGGGCA AGTGCTGAGC CATTGCTAAG ATTCCCCTTA CCCGTGCTTG	10980
TCCCTTGTAT TAAGGCACAA GGGCCCTTTG AAAAGAATTT TACCTGCTTT ATCAATTGAA	11040
AGGGATTAAG ACCTTGGGGG CCAACCCAAA ATAAACATGC GAACTTATTA TTTATAGGCT	11100
CCATGCACAC TTCGTAAAC CTCCATGGTC CTA CTGATTACCT CCACTCAATG	11160
AGAGGCAATT CATTACTGAA TGAGCCATAA GCGCCTCTTA TTTCGAGAGG GGGATGGCAG	11220
GA CTCAGTCG AGGAGAAGGA CCGCACCCAG GCAGCCTGGG CCCCTCGGCT CCTGTACTTA	11280
TTTACTGCTG GGTACTTCCT AGCCCAGCAT GTAATTACTG GTTCGTTT CAG TCATTTCGTTT	11340
AGTAAATGTT TCTTGGGCAC CTA CTACTACATA GGAGGCACAG GTCAAGGCAC TGGGGATATT	11400
CTTTCTACCC ACCCCCTCCC TCCCTACACT GTGATTAGGG ACTGACCGAT C	11451

22/33

(2) INFORMATION POUR LA SEQ ID NO: 3:

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
(A) LONGUEUR: 1834 paires de bases  
(B) TYPE: acide nucléique  
(C) NOMBRE DE BRINS: double  
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 3:

ATTTTTTTT TTTTTTTTGA GACGGAGTCT CACTCTGCCA CCCAGGCTGG AGTGCAATGG	60
CGCGATCTTG GCTCACTGCA ACCTCCGCCT CCCGGGTTCA AGTGATTCTT CTGCCTTAGC	120
CTCCTGAGTA GCTGAGACTA TAGGTGCCCC CCACCACGCC CAGCTAATTT TTGTATTTTT	180
ATTAGGACGG GGTTCACCA TATTGGCCAG GCTGGTCTCG AAATCCTGAC CTTGTGATCC	240
GGCCACCTCG GCCTCCCAA GTGCTGGGAT TACAGGTGTG AGCCATTGCG AGCAGCCCAG	300
AACTCAATTC TTAACCTTTA AAGTATGATG AGAAGAAGGA TCAAGCCCTC ACCAGCCCAT	360
TTAAGGAGTT TAGGCTCACT CTTGAGGATG TGAGAAGTCA TTGCTATTGG GTTTCACACT	420
GAGGTAAACA GGTGAAGTCA GCATTTTGGT AGTTCACAGC AGCTGCAACT CTTTGTATTT	480
CTCTGATACC TCCTGTCCCA ACCTACATCA GGCCTTCCCT TCTTCCTGCT TCCTTAATTC	540
CTCCATTTTC CCACCAGATG GAAGGACTGG AGCTTTGTGG ACAAAGATGA GAAGGCCCGT	600
CTGCAGCACC AGGTCACTGA GGATGGAGAG TTCTGGTGAG TCCAGAACCC AGGAAGACCC	660
AGAAGGGTAA GGGTGGGGAA GAGAGGGGAA ATCTCAGACC TCAGTCCCCA GCTAAGGTTA	720
TCAGATTCCA GCCCTTGGGA GATCTTGGCT GTGTTCTCCT CCAGCCCAAG GCCCAGCAAG	780
GATGAGGTTC TGAGAGGAGC CTTCCAGGCC ACAGGGACAA TGAGCCCAGG ACCAGGCCAA	840
CATGACATGG CTCTTGCCCTC CTGTGTGCCC CTCCGCCACA CACTCTATTC CAGCCACAGG	900
CACCCTGGCC TTAGCACAAT TCTTTTCTGA GCCTAGGAAG CTCCACTTAC CCTGATCTTC	960
CAACGTCAAC CTCACCCTCT CTCAGGTGTG TTCTATTAGG GCTTCAAGTC TCAGCTTAAG	1020
GAGAATTTTC AAGTCTCAGC TTAAGGAGAG CCCCCTAAGT TCCCCGAGGA CTGGGATTAA	1080
TTTATGATGC TCATCACCCT TAAAATTGTT TGCTTAAGCC GGGCGCGGTG GCTCACGCCT	1140
GTAATCCCAG CACTTTGGGA GGCCGAGGTG AACGGATCAC GAGGTCAGGA GATCGAGAAC	1200

FIG. 8C/1

SUBSTITUTE SHEET (RULE 26)

23/33

ATCTTGGCTA ACACGGTGAA ACCCTGTCTG TACTAAAAAT ACACAAAAAA AGTAGCCGGG	1260
CGTGGCAGCG TGGCCTGTA GTCCTAGCTG CTGGGGAGGC TGAGGCAGGA GAATCACTTG	1320
AACCTGGGAG GCAGAGGTTA CAGTGAGCCC AGATTGCGCC ACTGCACTCC AGCCTGGGCG	1380
ACAAGAGAGA CTCTGTCTTG GAAAAAAAAA AAAAAATGTG GTCTTAGTTT AATGTCAAGG	1440
GAAAGGTTTT GGGTGTTTTT ATTACTTTAT TTTTATTTA AAAACTATAA TAGAGACGGG	1500
CCTCGCTATA TTTCTCGGGC TGGTCTCAAA CTCCTGGGCT CAAGCGGTCC TCCCACCTTG	1560
GCCTCCCAAA ATGCTGGCAT GTGGGCCTGG TCAACATATG GGACCCCAAC TCTACAAAAA	1620
ATTTTAAAAAT TAGCCAGATG TGGTGGCGTG TGCCTGTAGT CCCAGCTACT TGGGAGGCTG	1680
AAGCAGGGGG TCACTTGAGC CCAGGAGGTT GAGGCTGCAG TGAAGTATGA TTGTCGTTCA	1740
CTTTTCTTCT GAACGTGAGA TTAAGTGTAG TCAGCAATTT GGCTTAGGAT TATTTATTCA	1800
GAATTTTAA CCGTCACGTT GCGGCAAACC AGGT	1834

FIG. 8C/2

SUBSTITUTE SHEET (RULE 26)

**24/33****(2) INFORMATION POUR LA SEQ ID NO: 4:****(i) CARACTERISTIQUES DE LA SEQUENCE:**

- (A) LONGUEUR: 14664 paires de bases
- (B) TYPE: acide nucléique
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

**(ii) TYPE DE MOLECULE: ADN (génomique)****(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 4:**

AGGAGGTGGA GGTTCAGTG AGCCAAGATC ATGCCACTGC ACTCTAGCCT GGGCAACAGA	60
GCGAGACTCT GTCTCAAAAA ATACACACAC ACACACACAC ACACACACAC ACACACACAC	120
ACACACATAT ATATACACAC ATATATATAC ACACACATAT ACACACACAC ACGTCTGTAT	180
ATATATGTGT GTGTGTATAT ATACACACAC ACACTATTCT ATATATTCTT GTAGAGCTAT	240
GTGTGTCTCC TGTGCTATTG AGCATGAGCC CTTTTTTTTT TTTTTTTTTT TTGAGACAGA	300
GTCTCACTTT GTCGCCCAGG CTGGCATACA ATGGCGCAAT ATCGGCTCAC TGCAACCTCC	360
GCCTCCTGGG TTCAAGTGAT TCTCCTGCCT CAGCCTCCCA AGTAACTAGG ATTACAAGTG	420
CCCGCCATAA TGCTCAGCTA ATTTTTGTAT TTTCAGTAGA GATGGGGTTT CACCATGTTG	480
GCCAAGCTGG TCTCAAATC CTAGCCTCAG GTGATCCACC TGCCTCAGCC TCCCAAAGTG	540
CTGGGATTAC AGGCATGAGC CACAGCACCC TGGTGAGCAC TAGAGCTTAT TTCTTCTATC	600
TAACTGTATT TTTGTATCCA TTAGCCACCC TCTTTTCATC CTCCCCTCTC CTTCCCTTCC	660
CAGCCTCTGG TAACCACTGT CTGCTCTCTA CTTCCATGAC ATATGCTTTG TTTTAGCTCT	720
CACATATGAG TGAGAGCATG CGACATTTAT CTTTCTGGCC CTGGCACATT TTTGAATCAT	780
TGTTAGAAAA GATGATGGTT TGGAGTAGAT ACATCAGAAG TGACAGCGTT TGCCCTAAAA	840
AGGAAAGACA GGCTCCTCTG GGACCCTGAC CAAGTTCCTG TGAATATTT TATTATTGTG	900
CTGTGTTAGT CCTGGGGTCT TCCGTCCCA GCCCTCCTCA CCTGCTCCCA TATGGCTCTC	960
TCTCTTCTTC CAACCTCTCA GGATGTCCTA TGAGGATTTT ATCTACCATT TCACAAAGTT	1020
GGAGATCTGC AACCTCACGG CCGATGCTCT GCAGTCTGAC AAGCTTCAGA CCTGGACAGT	1080
GTCTGTGAAC GAGGGCCGCT GGGTACGGGG TTGCTCTGCC GGAGGCTGCC GCAACTTCCC	1140
AGGTGGGAGA TGCTCTTGAT GGGGGGAGGG TCTAAGCCGA AAAAGTTCCA GGCAGAAGAA	1200

FIG. 8D / 1

**SUBSTITUTE SHEET (RULE 26)**

## 25/33

GCCTAACTAG TGCTTATTAA GTCTCTCTGT TCCAGACGTC CACTATCTTA TTAAACCTTC	1260
CCTGTTTTAC TGAGAAGGAA ACCACCATGC TGAGAAGTTT GCAATAGGGA GCTGGGTAGC	1320
AACTTTGGAA GCAGGAAGTT GTGGGAACAA TGCAGATGCT GCTTGACTT ACGATGAGGT	1380
TATGTCCAGA TAAGCCCATC CATCTTTTGA AAATACCCTA AGTGAAAAGT GCATCCAATA	1440
TGCCTAACCC CCCAAACCTC ATAGCTTACC CTGGCCTACC CTCAAACATT GCTCGGAACC	1500
CTTGACCTTA AGCCTAAAGT TGGGCCAAAT CATCTAACTC CAAAGCCTAT TTTACAAAGA	1560
AAGTTGTTGT AATATCTCCA TGTAAGTTAC TTAATACTTG TACCTAAAAA GTGAAAAACA	1620
AGAATGGTTG TACGGGTACT CGAAATCCAG TTTCTACTGA ATGTGCATCT CTTTCACATT	1680
GTAAAGTTAA AAAATTGTAG CCGAACCATC CTAAGTCAGG GACTGTGAGT ACTGTGTCAG	1740
TAACAGTAAG GGCATATTG GAGAACCAAG TTAGCAGCTG CTGCAATAGT TCAAGTCAGA	1800
GATGATGAAA ACCTAGACCA AGTCAGTAGC AGCAGAGATG GAGGGGAGAC AGCAGATTTA	1860
GGGAGAGCAT ATTGGGTGAT GTAGGGAAGG AAGAAGAATG ATGTCAAGAT TCCCAGTTGG	1920
GGACCTGACA ACATTGCAAC ATAAGACACA CAAGAAGATC GGGTGGGTGG CTCATGCCTA	1980
TAATCCCAGC ACTTTGGGAG GCAGAGCCAG GAGGATCACT TGAGCCCAGG AGTTCAAGAC	2040
CAGCACAGGC AACATAGTGA CACCTCATCG TTACCCAAAA TAAAAAAAAA AATGAGGTGG	2100
GAGGATTGCT TGAGCTCGGG AGGTTGAGGC TACAATAAAC TGTGATCATG CCACTGCACT	2160
CCTGCCTGGG TGACAGAGTG AGACCCTGCC TCAAAAAAAAAA AAGACACACA AGAGAAAAAT	2220
ATCAGCGTGT TGTTTGTTTT TGGTGGAGTT AATTGTGGGG TTCTAGGGAA AGGAATTTAG	2280
CTTGGGACAT GGAAAGTTTG AGGTTCCCTGT AGAGTGTCCT AGTGAAGATT TGTAATAGAG	2340
CATCGGATGC GCATATTAGA TGGCACTTGG TGATATGATA AGAACTCAAA AAATATTTGA	2400
GGAATAAAGG AAAGAAGAGG CCAGACGTGG TGGCTTATGC CTGTAATCCC AGCACTTTGG	2460
GAGGCTGAGG CAGGCGGATC ACTTGTGGTC AGGAGTTCGA GACCAGCTTG GCTAACATGG	2520
TGAAAACCCA TCTCTACTAA AGATACAAAA ATTAACCGGG GATGATGGTG GGTGCCTGTA	2580
ATCCCAGCTA CTTGGGAGGC TCAGTCAGAA GAATCGCTTG AACCAGGAG GCGGAGGCTG	2640
CAGTGAGCCG AGATCGCGCC ACTGCACTCT AGCCTGGGCA ACAGAGCCAG ACTCCGTCTC	2700
AAAAAAAAAA AAGTGAGAGA GATTGAGGCT GGGATATATG GCTCAGGCAT CATGCGCGTG	2760
TAGGGGGCAG TTAAAAAGCA GAAGTAAGAA AGATTGCCTA GGGAGGCAGG AAGGGTGAGG	2820

26/33

TGAGAGGAGA AGAGGCCAG GACCAGATTC TAGTCACCAA CAGCGTTTAA GGGGCAGGTA	2880
AGGAAAACAA AACCATCAGC AAAGACTGAG AATGAAAGCC CAGAGAGGAA GGAAAAGCCA	2940
CACATACAAT CAGTACAGCT CCATCTGAAT AAAGGTAGCG CCCCCCCCCC CCCAAATCAT	3000
TAGAGAAATG CCTGATTCCG TTTTCTGTGG ATTTTTCCTA AGAACCTAGA TGTGGGGAAT	3060
AGAAATAAAT GGTTCCTCT GTCTCATCCC CTCCCTGCCC TCTGAGAGGA AGCTGTGATT	3120
GCGTGCTCCC TTTCTGGGGG TGCAGATACT TTCTGGACCA ACCCTCAGTA CCGTCCGAAG	3180
CTCCTGGAGG AGGACGATGA CCCTGATGAC TCGGAGGTGA TTTGCAGCTT CCTGGTGGCC	3240
CTGATGCAGA AGAACCGGCG GAAGGACCGG AAGCTAGGGG CCACTCTCTT CACCATTGCC	3300
TTCGCCATCT ACGAGGTGTG TAGTCCTGAT TGGCTCCAGC CCAGGAAACA TACTTTCCCA	3360
GAGAGGACGC TTCCAGGGGC TTCTAGAGGG GCCCTCTGCT TCCTCAATAC CAGTGACCCA	3420
CAGAGCTCCT GGTATCAGGA CCACTTGTGT TTGTAACAAG CAAAAAATAC CAGGGGGGGC	3480
ATTAGAGAGG CAGTGGAGCG GGCCTGGCAG AACAGGTGCC TGGGGGTCAG GCTTCCGCAT	3540
GCGGGCTGCA GTTGCTGGCA TTGCCTTCCG CAGGCTCCTC ATCCTCATTG ACATCTGAAG	3600
CATCTTCCTT TCTGTTTCTT CTCAAGGTTT CCAAAGAGGT ATAGCAGCAG CACCGGCCAG	3660
CAGTTGTGTG CAGCACTACC CAGGGGGGCC CGACTCTGTC TGTGGCTCGT CGAGAAGCTT	3720
CCTGGTGGGG TTTGTGGGCA GGACTTGTGA TAGGAGAGGG CTTGCCTGT TGTTATTTC	3780
CACTTGCAGA GCAGGTTGCC TCAGGGCATT GCATGACCCA TGA CTACCAC CCCCAGGATG	3840
TGCACTTTCT CCCTCGCACC AGACACTGCA CGTCACACAC ATGCCTTTGC AACTCACCC	3900
TCCTCCACGC TTACAGCCAC ACACACAGTC ACACAGACGC GTTCTGAGGG TGGCTGCCCC	3960
CTTGGGATGG AGGAATCACT TCCCTCAGAA CCCAGCCAAG TCCTCTAGGC CTCCTTGGGG	4020
GTCCTTCAG CCTGAGGGGC TTCGGAGCTG AGGACAGCTG TTCTGGTAAG TGTCCCTGAG	4080
TGTGGGGATG ACACATTTCC ATTCACTCTG AATCACAACA GAAAAGGGAA GAGGAATTGA	4140
GGTAGGGAGC CTATTTAACC CTTGGGAGTC GGAAGTAGG GAGGTTGAAA CTGTGACATG	4200
GGTGACCAGG GAGTTGGGAA GGGACCCTTG GAGGTGGCTG TGGCAGGACA GGACGTTCTT	4260
CCCAGGGGGC TCATGTGCC TGGGCTCTCC CCATCTCTCA GATGCACGGG AACAAGCAGC	4320
ACCTGCAGAA GGA CTCTTCTT CTGTACAACG CCTCCAAGGC CAGGAGCAAA ACCTACATCA	4380
ACATGCGGGA GGTGTCCCAG CGCTTCCGCC TGCCTCCCAG CGAGTACGTC ATCGTGCCCT	4440

FIG. 8D/3

SUBSTITUTE SHEET (RULE 26)



27/33

CCACCTACGA GCCCCACCAG GAGGGGGAAT TCATCCTCCG GGTCTTCTCT GAAAAGAGGA	4500
ACCTCTCTGA GTGAGTGCTG GCCCAGCTTT CCCACGTGTT TCTAAAAGCT CACATGGCCC	4560
ACTCCAGAGG TTGAAGGCAT GAGGCAGCTA GACACGTCTC CTCCAGGGTC CTTCTGCTGC	4620
TCCTGAGCCA CTGGCCACAT TACCCCCATT CATTCAATTCA TCCATTCTGT GATATTTATT	4680
GAGCACCTAC TATGTTCCAG GCACTGTCCT AGGCACTAAG GATAGAGTAG TGAAGTAAAC	4740
AGAAAGAAAT CCCTGCCTTC ATGGAGCTTA ATATTCTAAC ATGAGACAAT AATGGATAGG	4800
AAAAACATAT GTAGCATGTT AGATTTGGAG AGGTGATATG GAGCAAAAAT AAAGTAGGGA	4860
AGAGGGATAG GAGGTGTTGG GGATGCTTGA AATTTTAGGT TAGCATGGCC AGGAAAGCCA	4920
CATCCTGTCC CTGGCCACCA CAGATGAGCT CATAGCCCCT GCCACTCTGA TCTCTGTCCT	4980
TGGAAGATGC ACCAGGTCCA TGGGTAGGTG GCTGGGTCAT GCCTTTGGGG GGCTCTGAGC	5040
AATACTAACA AGAACCTGCG TGCCTGGGCT TGGCTGTCCG GGATGGTGCT GACATGGGGC	5100
TGTTTCCTGG GGTGGGGTG TTCCAGGGGT TCTCTAGAGG CTGGTTCTGG CTTGGCTGCC	5160
AGGAAGCCGT GCACCAGAGC AAACCGTCCA CGGGCCTCCT GCTTGCTTCT GGTGACACTG	5220
AGACCCACA TGTCTGTATT CCTCACAGGG AAGTTGAAAA TACCATCTCC GTGGATCGGC	5280
CAGTGGTGAG TGGTTTAGAT CTTCTGTGCG AAAAGTCCAG AGGGTCCCCT TCCCTGACCA	5340
TGCAGGGGAC AGATGGTGCA GGGGAGAATG GGCAC TGGA GAGGGAATGG GAGTCTGGGC	5400
TGTGCTGAGC AGTCCCTCCT TGGCACTGCA AATCCTACTT TGGCATGGCC AGAAGTAATC	5460
GGCCTTAAGC ACCGGGGGCC ATTGAGGCAG TTCAGGGGCT GGGAAATATG GAAGAGGGTC	5520
CTGGAAAGGA GAAGCAATTT GAACAATCGG AGGGAACAAG GCCACAGGAA GGGATGACAA	5580
GAGCCGCAGC GAACACTGGA TTCTGAGACT GGATAACATT GGATTTTACA CATAGAGAAA	5640
AGAAAGTAAG CTGGTGCCGG ACCTGGTGTT GACACTTGGA TCCTCCACTT ACCAGCGGGG	5700
TGACCTGGAC AATTTCTGTA ATCCCTCTCA CTCAGTTTCC TACTCAGTAA AACGGGGATG	5760
ATAATGTGCC TTGCAAGGCT TTTGTGAGGC TTCATCAATG AGGTGATGTA TGTGAAGTGT	5820
CTGGCACAGC ATGGGCACTC AAACAGAGGT GCTTTTTTAC ACTTTACACC TTACAAGGTA	5880
CTTTTCACAT GTGTCATCGC GATACTTGCA AGGTTGCTGA GAGGTAGATG GGGTTATAAT	5940
CCCTGGTGTT CAAGAAAGGA AGCAGAGGCT CAATGGGGTT GAATGACTTC TCTGAGTTCA	6000
CAGAGCTCAG TAACTGGCAG GGTTTGGAAC TCACATTCAG ACTCTCTGAC TCCAGACTTA	6060

FIG. 8D/4

SUBSTITUTE SHEET (RULE 26)

28/33

GGTTTTTCCG CACCTCCACG CTGAGGCCAG CCCCAGGCAG TGAGAAGCCC AAAGTCCGAA	6120
GCACAGAGTG CTGTGTGTTG GGCTCTGTGT GTTGAGGAGT CTTGTGACTG CCTTGGGGCT	6180
TTGGGCTGTA GTCAGCTGAC AGTCCTTTGT GCTCTGTGGG GATGACGTAG GCCAATGGGA	6240
GGACAAATGC CCCTCTGAAC TGTCTTCTGG GCAGTGACAG TCATGGTCAT AATCCTGACC	6300
CTGAGCCAGT GCCAGGTCTC CAAGTGCCTT CTGAATGACC ACAGGCGATT GGTTTTAGTG	6360
GTAGGTGCGT GGGGATCTGT TCTGGTCATC TGGATGCTGG TCATCGGGTG CAGTATTGAT	6420
CAGGACCTGC AAACCCAAAA GCTTATGGGA GCTGGCACGT CACGTGAGTA GAGCAGGCAG	6480
GTGCAGGGTT TTTGATGTCC CTGCACTGAC ACAGTTGTCT GCAGTTCTCC AATTGACAT	6540
TTGGGCTCCA GTGTGAGGG TCAAACAAGG AATTTTGGGG CGTGGGCCAA ATCTGGGAAG	6600
ACACAGGGAG CAGGGCCCTT TGGCTCAAGC TGATAGTTGC CGCAGGGATT ACCAGGCCCA	6660
GGGCAGCCTG CCACAAGCTG GGGCTTTTAC CAAAGAAAAT CTCCCTATGT TAAATGCTTG	6720
CTCAAAAATT TTTAAAAAAT ATTCTGTAAG TCAAAATCCA TTGTTAGGTC AGTTTGAGAG	6780
AGCCATGTTT TTGGTGTTTT AGTAACCAAT TTCATTTTTT TATTATTAT TTATTGTTT	6840
ATTTTTGAGA CGGAGTTTCA CTCTTGTAC CCAGGCTGGA GTGCAATGCC ATGATCTCAG	6900
CTCACTGCAA CCTCCGCCTC CCGGGTTCAA GCAATTCTCC TGCCTCAGCC TCCTGAGTAG	6960
CTGAGATTAC AGGTGCCCAC CATCACGCCT GGATAATTTT TGTATTTTTT AGTCGAGATG	7020
GGGTTTCACC ATGTTGGCCA GGATAGTCCT GAACTACTGA CCTCAGATAA TCCGCCCACC	7080
TCAGCCTCCC AAAGTGCTGG GATTACAGGC ATGAGCCAGC ACGCCCGGCC ACCAATTTC	7140
TTTTTTAAAA AAGGAAGAAA GAAAACCTTA GCCAGAAGAT CTTTTCTCTT GCCATATGCA	7200
GTAAGAGTAG ATTATAAAAA CAAAGTCAGA GCAGTCACTG GTGTCTGGGC ATGGAGGAGA	7260
AAGAAGAATT CTCTTCTCCC TTCACCCTCC ATGCCCTTTT TTGGCTCCAT GTGATTGAGA	7320
TTTCTGGACC CTGGAGCCCC ACCCCAAGCT AAAGACCAGG ATACAGGGAA GCCACAACCA	7380
CTGGCGGTTT TGAGAACTTA CTTTTCACTT ATTCTGCATT TACTGTTTCC TTTTCTTATG	7440
CAGAAAAAGA AAAAAACCAA GGTAGGTGTG TGGGTAGAGA GCATGAAGTG TGTGACTCA	7500
TGCATATGTA TGTGCATGCA TGTGAAGTGT GCATGTGTGA GCTCATATGC ATCCATGCAC	7560
CAGACTTGCC TCTTCCTCCC CCTCCTTCCT GAGCTTCTGC TGGGGCCGAG CGTGCAGTAA	7620
TGACAACTAC GATTTGCTGG GGAAGGCTA CGTGCCAAGC ACTCTTTTAG GTGCTTTCCA	7680

29/33

TGATTAATTC CTTCTCACA ACAGCCCTAT GAGATTAGTA CTATAACTAT CCCCATTTTC	7740
AGAGGGAGAA AAGGTACAGA CTTGACTAAC TTGCCCAAGG CCACACAGCC AGAGAGGGGC	7800
AGAGCCAGTA CTTAGAGCCA GGCAGTCTGG GTCCAGAGTC CGTGTCTGA ACCACAAGAG	7860
GCCATCATAC GCCATCAGAT TTGGTGCTAG CATTCTGGT GGTGCCTGGT GGTGATGGAT	7920
CCATCACAGG GGTCTCCAG GTACTGGTGC TGGCCAGAC CAGAGCTGAC ACTCCTCAGG	7980
CACTACCACA TTCCAGGCAC TGTGCTTGGG GTCAGTCCCT CTCTTTTTTT TCCCCCCAA	8040
TTATAACAGT ATCTACAAAG TAGGTGCTGT TATTTTCCC CTTTCACAGG TGAGATAGAC	8100
TCAAAGAAGT GAACTTGCCC AAGGAACAGA ACTAATGAGT GGGGAAAATG GAACTGGAAA	8160
CCATGTCTGT TTAATCCAAA ACCTGTGTTT CTGCCCCTCT TTCTCTGATG CCAGCCCCCT	8220
ACACTTCAAG GCCTGTGTTG TCCAGACCCA CACTCGGGCC TGCCAGTGTG TGCCTGGCAG	8280
GGATGCTCCA TGGCCACACC ATATCCATCC TACACATCCC CCCTCAGACT GTGACCTCCA	8340
TTTGCTCTGG GATCCCCACA AGCTTCAGCT GCTTGAGCAA GACACTGCTT AGAAGGCAGA	8400
GCAAGCCAAG GCCTCTGGGG CCTGCTGGGA GCCAAAGCTG GGGAGCCGTT TCCAGGGGTC	8460
TATCTGCTTG AGCTGTCCTA GATGAGCAGC ATGGAAGGCC AGTGGTGCAT GAGTCCAGGC	8520
GGGCTGCTTT TCTGCTCCGA GAGGCTCTGC CTGCCAGTT GTTCTCTGCA TTGCAGCCTC	8580
AATCCCCACA GCCTTGCCCT CCCCCGGCTT TCCCTACAGG TGCACCGCAT CCACAGTGTT	8640
GGCACCATGC AGCAGCCGCT CTCCGTCCTT TTCATATCCT TGTCACCTGC ACGAGCATGT	8700
CTTGAAAATA TCCCTTGTTT GTGTAGCATC TTAAATGTTT TTGCAGTATG ATTTTGCATT	8760
CAGTATCTCA TTTGATCCCC ACAAGAGCCC TATGAGGAGG GAAAGCAGAT TTTACCATT	8820
AAGGATGAGT AAAGTGAGGC CAGAGAGGAT ATTTTGGTT TTTTTGAGA CAGTCTCACT	8880
CTGTCACCCA GCCTGGAGTG CAGTGGCTTG ATCTTGGCTC ACTGCAAGCT CCACCTCCCA	8940
TGTTACACACC ATTTTCCTGC CTCAGCCTCC CAAGTAGCTG GGACTACAGG CACCCACCAC	9000
CACACCCAGC TAATTTTTTT GTATCTTTAG TAGAGATGGG GTTTCACCCA GTTAGCCAGG	9060
ATGGTCTTGA TCTCCTGACC TTGTGATCTG CCTGCTTCGG CCTCCTAAAG TGCTGGGATT	9120
ACAGGCGTGA ACCCCCCTGC CCGGCCAGAG AGGATATTTC TTAATGAGGG GCAGGGCTGG	9180
GATTCCAGCC CAGTGTCTG ATGGCTCACC CACTGACCAT TCCACTAATC CGTGTCTTT	9240
TTCAATCTAA ACTTTCAGGG TTGTAGAGGT TCCTTTGAGG TGCCTCAGTA CTTCCATGGT	9300

FIG. 8D/6

SUBSTITUTE SHEET (RULE 26)

30/33

GATGTGGGGT CTGAGGGCCA AGAGCTCTGT TCTCATTAAAT CAGAGAAGCT TGTGTTTTTA	9360
AAAACACCAT GTTTACTGCA GGAAATTIAA TTGGACAGTG TTTCCATCTG GAAAAAAAAA	9420
AGTCTACAAA ATACTTGACA ATCACTGCAC TAGATCATGC TGCTTTTAGC ATTCTTAGCA	9480
TTTCACGTGC TGAGCTCTCA ATACTCTACC ATGAGGAGGG ATGGAGTGGG TATGAAAAGA	9540
TAAAGAACTG AAGTCACACG GCTTGTCAGT GGCAGAGATA GAGCTTGAAC CGAGGTTGAA	9600
GAGCTCCCGC CTATTCCTTT CCTCTTCTCA CTGGATAAAG CTGCTCCAAG AGAGGTGCTG	9660
CCTCAGTGTG CCTGTTTACA CTGTAATCCT CCCTTCCTTC CTGCCTCCTC CCTCCTCTCT	9720
CCAGCCCATC ATCTTCGTTT CGGACAGAGC AAACAGCAAC AAGGAGCTGG GTGTGGACCA	9780
GGAGTCAGAG GAGGGCAAAG GCAAAACAAG CCCTGATAAG CAAAAGCAGT CCCCACAGGT	9840
GTCTGGGCAT GTGGCATGGG TGGGGTGGCC AGCACGCTAC AGGGGCTTCC TATGCGCTTG	9900
GGATACACAG GGGCTGGAGG CTTCCCAGGA GTTTGTCTTG AACATCTGGA GGTTTGAATT	9960
TGTCCCACTG ACCTTTTCTT TCAGCAAGTT CCCCTGAAAT TTGGGCTGCT GCTTGGGTGA	10020
ATATCCCAGG ATGGGGGTTC CATTCTAGGA GTGGACTGGC AGGCTGAGCC TCCCATGGAG	10080
CTGATCCAGC CAGGATACAG AGAAGGGGAG GCAAAGGCTG AGACAGAACC AGCTTGAGAG	10140
CGGAGGCGCA ACTCTTGTCT CCTGGTGGCC TTGAGCATT CACAATAGGG GGATAAAGGA	10200
TAGGAGCAGA AAAGTGGGGC TGAATTCAGA AATGGGGTCC TCTAGAGCTC ACGGGAGGGT	10260
GTTAGATTGG AGTGGGAGCT TAGTGGAGGT GAGCCTTAGA GGCAAAGTC TCCAGACCAA	10320
TCCAGGCCCC CTCTTCTATC CGGGGGCCCC TCTTCTATCC AGGGCCCCTC TTCTGTCTGG	10380
GAGCCCCTCT TCTATCTGGG GCCTCATGCA GTGGGGCCTA GGGGAGGTTT TCTGAGGACT	10440
TGGCCTTGAT GACAGGGTGG CTGGAGGAAT CAGAACGGTC AGACCTTCTT TGACCTGCGG	10500
GCACCTTTAG TTGGAATGCT CAGGCCTGGG ATGGTGGAGG GGGCTCTTGC AGGTGGGGAC	10560
TGGGGTGGCG GGGAGGAGGC TGTATGGCCG CCATATCTCC TTTGGCTGGG GGCCTCAGGG	10620
CTGGAGAGGT GTGAAGAGTC CCTGAGGCCT CGATGCATCT CACTCCAGCT CACCAGGTCT	10680
GCATTTGCCC GTCCCCAGCT CCTGCTGCCA CCCCCGGCCG TTTTAGGCAC TTGGCTCCCT	10740
TGGCCCAGAG GAGCTTGCTT CACAGGCCTG TGCACCTCTG ACCCCTGTGA ACCAGTTTTC	10800
CTTTGTGCCT CCACAGCCAC AGCCTGGCAA CTCTGATCAG GAAAGTGAGG AACAGCAACA	10860
ATTCCGGAAC ATTTTCAAGC AGATAGCAGG AGATGTGAGT ACCTCCAAGC CCAGGACGCC	10920

FIG. 8D/7

SUBSTITUTE SHEET (RULE 26)

31/33

CACAGGTGCT TCCTTCTCTC CTGGATTAAC TGCTCAGATT ACCAATTATT TCATTATTGT	10980
TTGGTAGAGG TCACTTTTGA CTTCGGTGGA GCCAGGGGAT GTGTGCGTAG CACACAAATC	11040
CACAAGCCCT TGAGTTTTGG ACTGCCACGT CTGCTGGGGG GCTCAGAGGC CTTTTTGCTC	11100
TGAGCTGCCC ACGGTGGTCC TGATAGCTGA GGTGCAGTAT CTGGCCCCCT GTCTTCCTCA	11160
GAAAAGCCCC AGCTTCCCAT GACATAATAG CACCGACAGG GATTTTACAA ACACAGCCAG	11220
GTGGAATTTG TTTTGCAAAG TGTCCGCGCC AGGAGCTGCT GTACTCCTGA ACCATGACCC	11280
TCCTCTCCCT TCCTCCTCAG GACATGGAGA TCTGTGCAGA TGAGCTCAAG AAGGTCCTTA	11340
ACACAGTCGT GAACAAACGT GAGTTGCTCA AACCAAATGG GGGTGGGGTG GGTGGGGAGT	11400
CCCGTTGTCT CAAAGCAGCT CCTCACTCTT CTCCATCCCC CCAGACAAGG ACCTGAAGAC	11460
ACACGGGTTT CACTGGAGT CCTGCCGTAG CATGATTGCG CTCATGGATG TATCCTTCCT	11520
GCCGCCCCCT CCCGACCCTC TGTCATCAGC CCACGGGGGC CAAGGCAACA TACAGGGTGC	11580
CCAGTCAGGC AAAGGGCCCT AATTTGTGCC CAGGGAAACT TAAGGAGACC CTGATTCAGA	11640
ACATCTTGGA TACTCGTCTG AAAGGGGTTG TTAGAGGCGG AAGGGGAGGA TGTTGGGTTG	11700
TAACTGCCCT AACCCCTGTG CTTCTCTCAG GCCTGGGATC CTGCCCAAGC AAAAGTGGTC	11760
CTTAGGAGAG CGGCTCCTGG GTTACAGAGT AGGCGCAATC TCTGACTGGT GGTGGAGTGG	11820
AGGGGAGGGT TAAATAGTAC AACAGGGCAG TGGGTAGGAC AGCCCGGAGT CTCCTAGACC	11880
CTCCCTCCAA ATCCAGGGGG ATTTGCTGT GTGCTGTGTA GCCCTGACCT CCCTCCTCCA	11940
GACAGATGGC TCTGGAAAGC TCAACCTGCA GGAGTTCCAC CACCTCTGGA ACAAGATTAA	12000
GGCCTGGCAG GTGGGAAGAG AAAATGAAGC GTGGGAGTCA AGAATGGGGT TGATTTGGAG	12060
ATTCAGTGTG TGACCTCCAT CCTCAAATTT TCTATTGCCA GAAAATTTTC AAACACTATG	12120
ACACAGACCA GTCCGGCACC ATCAACAGCT ACGAGATGCG AAATGCAGTC AACGACGCAG	12180
GTGCTGAGAA GGAAGGGGTG TCAGGGATGT GGACCCGAGA CGGTGGGAGC AGGAATGGGA	12240
GGGGACTAGC TACTAGGGCC CCACTAGAGA AGGAGAGGGA AAGGGCTTCT CACTTTCCT	12300
TCCCAGGTCA CAGAGTGTCC GAGAGGCAGG GAAAATAGAA GACAGGCCCA AGGCCTCCAG	12360
CTCCACGTCC ACCTCTAACA TGGTCCCCTC CACAGGATTC CACCTCAACA ACCAGCTCTA	12420
TGACATCATT ACCATGCGGT ACGCAGACAA ACACATGAAC ATCGACTTTG ACAGTTTCAT	12480
CTGCTGCTTC GTTAGGCTGG AGGGCATGTT CAGTAAGTGG GAGAGGGGGG CTGCCCTCTG	12540

FIG. 8D/8

SUBSTITUTE SHEET (RULE 26)

32/33

CTCTCTTGCA GGGGCAGTTG TGGCAACAGG CATCTCACCT GATAATCTCC AGTCTGCTCC	12600
ATCCAGGCTG AACAAAGGGCC AATGACCTCT TTAGGCCCAG AATGGGATGG CAAAGGGAGG	12660
GTTACTGGTG ATTCTCTGCC TGCACATCTT TGTGCTGATG AGGGACAGCA CTGGGCACAC	12720
GGTCCTCTGA GGGGAAGTTA CAGTAGTAGA GGCGGAGTGC GCCTGTAACCT GGCCTCTGGC	12780
CTGTGCATTG TTTCACAGGA GCTTCTCATG CATTTGACAA GGATGGAGAT GGTATCATCA	12840
AGCTCAACGT TCTGGAGGTA AAGCATAGGC ACAGCACATT CCCCCTACAC ATTAAAACTC	12900
AAGGTGGAGG GGTCAACGGG GCGGACTGGA CCCAGGGTGT GCTCCTCATT TCCACACAGT	12960
GGTGGAGGGA AGGGATAGGA ACAGAACATG GAGGGAGGCT CAGCAGGCTC CCAGGACACA	13020
TGCACTTGAG GCCCAAAGG ACCTCTGCTC CCCAGTCAC TTGATGCGGG AAAACATGCA	13080
CCTTCTTAGG GAAGATCTAG GAGAAAGGAA ACAGTAAGCC ACTGCTTCTT GGAAAATCTT	13140
CTGGGGGTCT GACCTGCTGG GACTGTTCCC TTTCTCTTG CCCCCTAAGA TTCCTAGGGC	13200
GGGGGGGGG GGGGCTCACT CTTTTCTGAT CTACATTCTG ATCTTGGGAC TTCTTTCAGT	13260
GGCTGCAGCT CACCATGTAT GCCTGAACCA GGCTGGCCTC ATCCAAAGCC ATGCAGGATC	13320
ACTCAGGATT TCAGTTTCAC CCTCTATTTT CAAAGCCATT TACCTCAAAG GACCCAGCAG	13380
CTACACCCCT ACAGGCTTCC AGGCACCTCA TCAGTCATGT TCCTCCTCCA TTTTACCCCC	13440
TACCCATCCT TGATCGGTCA TGCCTAGCCT GACCCTTTAG TAAAGCAATG AGGTAGGAAG	13500
AACAAACCCT TGTCCCTTTG CCATGTGGAG GAAAGTGCCT GCCTCTGGTC CGAGCCGCCT	13560
CGGTTCTGAA GCGAGTGCTC CTGCTTACCT TGCTCTAGGC TGTCTGCAGA AGCACCTGCC	13620
GGTGGCACTC AGCACCTCCT TGTGCTAGAG CCCTCCATCA CCTTCACGCT GTCCCACCAT	13680
GGGCCAGGAA CCAAACCAGC ACTGGGTTCT ACTGCTGTGG GGTAAACTAA CTCAGTGGA	13740
TAGGGCTGGT TACTTTGGGC TGTCCAATC ATAAGTTTGG CTGCATTTTG AAAAAAGCTG	13800
ATCTAAATAA AGGCATGTGT ATGGCTGGTC CCCTTGIGTT TTGTTGTCTC ACATTTAGAT	13860
ATCAGCCATG CATGACTGAA TGGCTTCCAA TCATATACTC ACCTATCACC TACAAGAGAA	13920
CAATGAAAAA CACACACAAA AACAAAATCT TGAATTTTGT AATCATGCCT ATTGCTATTT	13980
CTTGAGCATA AGAATGGCTC AGATACTTTC CAAGACATAA AAGGAAGGCA GAGGAATAGT	14040
TGTTGCTGTA AAAGACATCA AGAATAAATG GGGTCATGTA CAACGGGAGG GGCCGGTTAC	14100
CTGAATAATG GAGTGGAGAT TGAGCTATCC TAGCTCCTCT GCTCACTAAC TGACCTGTCC	14160

FIG. 8D/9

SUBSTITUTE SHEET (RULE 26)

33/33

CATGACCGTG GACAAAACCC TGAACGCAGC TGTGTTGTTG CTAAACTTCT CTGGACCATG	14220
GCCTGCGGCA TATCTATAGG CATCCTGTGT TTTCCACCCA GTTTCCTTCT TCCTCGCTAA	14280
GCCAACGTGG AAAGGGCTGG CCGTGAATAT GCAGACAAGG TAACGAAAGT AAACCGTCAA	14340
TTAGTAAAAG TACTTCATTT TCCTCTTGTA TTGCTTCAT TCTTGCTTCA CAAAGTTACG	14400
AAGTCCACAG CTTTATACCA AAATGTAAGA AGGCTATTTG CTTATAAACA TTTTGAGTCA	14460
GGTGTCACTT GATTTCATTC TTCTAATCCA TATTCAATAT TAAAAAATCA GAAACCAAGG	14520
GTGCTGGAGC AGCTCTAGGG CATATATTTT TCTTAAATAG GAGAAAGATT TTCAACAGCT	14580
TTTCCTCCTT GACCCCCTCC TTCCCAATT TATTTGGGTC ACTACCTTGA ATTTAGAGTG	14640
AATCTGGGAA ATGTAGTCAC CAGG	14664

FIG. 8D/10

SUBSTITUTE SHEET (RULE 26)

## PATENT COOPERATION TREATY

PCT

## NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

United States Patent and Trademark  
Office  
(Box PCT)  
Washington D.C. 20231  
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year)

23 July 1996 (23.07.96)

International application No.

PCT/EP95/04575

Applicant's or agent's file reference

B2628A - FL

International filing date (day/month/year)

21 November 1995 (21.11.95)

Priority date (day/month/year)

22 November 1994 (22.11.94)

Applicant

BECKMANN, Jacques et al

1. The designated Office is hereby notified of its election made:



in the demand filed with the International Preliminary Examining Authority on:

19 June 1996 (19.06.96)



in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was

was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Mirjam Van Straten

Telephone No.: (41-22) 730.91.11



## PATENT COOPERATION TREATY

PCT

COMMUNICATION OF  
INTERNATIONAL APPLICATIONS

(PCT Article 20)

From the INTERNATIONAL BUREAU

To:

United States Patent and Trademark  
Office  
(Box PCT)  
Crystal Plaza 2  
Washington, DC 20231  
ETATS-UNIS D'AMERIQUE

Date of mailing:

10 October 1996 (10.10.96)

in its capacity as designated Office

The International Bureau transmits herewith copies of the international applications having the following international application numbers and international publication numbers:

International application no.:

PCT/EP95/04575

International publication no.:

WO96/16175

**CORRECTED VERSION  
VERSION CORRIGEE**The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer:

J. Zahra

Telephone No.: (41-22) 730.91.11